

RA Xu S.-H., Gu J.-X., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,
RA Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.,
RT "Gene expression profiling in the human hypothalamus-pituitary-adrenal
RT axis and full-length cDNA cloning.",
RL Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548(2000).
RN [6]
RP SEQUENCE OF 232-759 FROM N.A.
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.X., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.,
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.",
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC TISSUE=Colon, and Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X.J., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickinson W.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. This cytoskeletal protein co-
CC localizes with actin stress fibers and focal adhesion plaques.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing. Named isoforms=3;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Names=Beta;
CC IsoId=Q9UHB6-1; Sequence=VSP_003116;
CC Name=Alpha;
CC IsoId=Q9UHB6-2; Sequence=VSP_003116;
CC Name=3;
CC IsoId=Q9UHB6-3; Sequence=VSP_003117;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta, kidney,
CC pancreas, prostate, ovary, spleen and heart. Also detected in
CC lung, liver, brain, skeletal muscle, thymus, testis and intestine.
CC Not detected in leukocytes. Eplln-beta levels are generally very
CC low. Eplln-alpha is abundant in epithelial cells from mammary
CC gland, prostate and in normal oral keratinocytes. Levels are low
CC in aortic endothelial cells and dermal fibroblasts, and not
CC detectable in myocardium.
CC -!- INDUCTION: Down-regulated in some cancer cell lines. Eplln-alpha
CC is induced by serum. Eplln-beta is constitutively expressed.
CC -!- SIMILARITY: Contains 1 LIM zinc-binding domain.
CC -!- CAUTION: Ref.4 (BAA91120) sequence differs from that shown due to
CC a frameshift in position 697.
CC -!- CAUTION: Ref.6 sequence differs from that shown due to frameshifts
CC in positions 365 and 662.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; AF198454; AAF23755.1; -
CC EMBL; AF198455; AAF23756.1; -
CC EMBL; AL136911; CAB66845.1; -
CC EMBL; AK000372; BAA91120.1; ALT_FRAME.
CC EMBL; AK000335; BAA91092.1; -
CC EMBL; AK023649; BAB14625.1; -
CC EMBL; AK000057; BAA90914.1; -
CC EMBL; AF157325; AAF67491.1; -
CC EMBL; AF218025; AAG17267.1; ALT_FRAME.
CC EMBL; BC001247; AAH01247.1; -
CC EMBL; BC010664; AAH10664.1; -
CC HSSP; P12965; ICTL.
CC InterPro: IPR001781; LIM.
CC Pfam: PF00412; LIM; 1.
CC ProDom: PD000094; LIM; 1.
CC SMART; SM00132; LIM; 1.
CC PROSITE; PS00478; LIM_DOMAIN 1; 1.
CC PROSITE; PS0023; LIM_DOMAIN 2; 1.
CC Cytoskeleton; LIM domain; Metal-binding; Zinc; Alternative splicing.
CC DOMAIN 398 448
CC VARSPLIC 1 160
CC VARSPLIC 1 302
CC CONFLICT 344 344
CC CONFLICT 381 381
CC CONFLICT 415 415
CC CONFLICT 463 463
CC CONFLICT 491 491
CC CONFLICT 520 521
CC SEQUENCE 759 AA; 85225 MW; 996378AFD3B003D5 CRC64;
Query Match 100.0%; Score 3927; DB 1; Length 759;
Best Local Similarity 100.0%; Pred. No. 2.3e-178;
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MESSPFNRQWTSLSRVTAKELSLVNKKSAIVEIFSKYQKAAEETNMKKSNNTENL 60
DB 1 MESSPFNRQWTSLSRVTAKELSLVNKKSAIVEIFSKYQKAAEETNMKKSNNTENL 60
QY 61 SQHFRKGTLTVLKKKNWNPGLGASHSTSLRNSSTSLRHRADHPPEVTSAAASAKAQD 120
DB 61 SQHFRKGTLTVLKKKNWNPGLGASHSTSLRNSSTSLRHRADHPPEVTSAAASAKAQD 120
QY 121 EQIHPRSRRLSPPEALVQGRYPHIKQGEDLKDHSSTESKQENCLGESRHEVEKSEISEN 180
DB 121 EQIHPRSRRLSPPEALVQGRYPHIKQGEDLKDHSSTESKQENCLGESRHEVEKSEISEN 180
QY 191 TDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRASAGRKISSENSYSLDDLEIGPGQ 240
DB 191 TDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRASAGRKISSENSYSLDDLEIGPGQ 240
QY 241 LSSSTFDSEKNSRRNLRLPRLSETSKDMAKYQAAVSKOSSSTNTYNELKASGGGIKI 300
DB 241 LSSSTFDSEKNSRRNLRLPRLSETSKDMAKYQAAVSKOSSSTNTYNELKASGGGIKI 300
QY 301 HKMQCKENVDPGEVCITHQEGEKISANENSLAVRSTPAEDSDSDSQVSKSEVQVHPKP 360
DB 301 HKMQCKENVDPGEVCITHQEGEKISANENSLAVRSTPAEDSDSDSQVSKSEVQVHPKP 360
QY 361 LSPDSRASSISESSPPKAMKKFOAPARETCVEQKTYPMERLLANQQVFHISCFRCSYC 420
DB 361 LSPDSRASSISESSPPKAMKKFOAPARETCVEQKTYPMERLLANQQVFHISCFRCSYC 420
QY 421 NNKLSLGTASLHGRYICKPHFNQLFKSKNGYDEGFGHREPHKDLWASKNEBEILERPAQ 480
DB 421 NNKLSLGTASLHGRYICKPHFNQLFKSKNGYDEGFGHREPHKDLWASKNEBEILERPAQ 480
QY 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASSQSQEKDKPAETKKLRITAWPPPTELGS 540
DB 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASSQSQEKDKPAETKKLRITAWPPPTELGS 540

```
QY 541 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLKLRLSSSLKRSRPTTVAASFOST 600
Db 541 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLKLRLSSSLKRSRPTTVAASFOST 600
QY 601 SVKSPKTVSPPIRKWMSQSESVGGRVAERKQVENAKASKKNGVNGVKTITWQNKESKG 660
Db 601 SVKSPKTVSPPIRKWMSQSESVGGRVAERKQVENAKASKKNGVNGVKTITWQNKESKG 660
QY 661 ETGKRSEKHSLENNENLVENGADSDDEDDNSFLKQSQPOEPKSLNWSFVDNTFAEFT 720
Db 661 ETGKRSEKHSLENNENLVENGADSDDEDDNSFLKQSQPOEPKSLNWSFVDNTFAEFT 720
QY 721 TONQKSDQVELWEGEVVKELSVBEQIKRNYDEDEDEE 759
Db 721 TONQKSDQVELWEGEVVKELSVBEQIKRNYDEDEDEE 759

RESULT 2
EPLI MOUSE
ID EPLI MOUSE STANDARD; PRT: 753 AA.
AC Q9ERG0; Q9ERG1;
JT 28-FEB-2003 (Rel. 41, Created)
JT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Epithelial protein lost in neoplasm (mEPLIN).
ZN EPLIN.
DS Mus musculus (Mouse).
JC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
DX NCBI_TaxID=10090;
ZN [1]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
RX MEDLINE=21100452; PubMed=11179679;
RA Maul R.S., Sachi Gerbin C., Chang D.D.;
RT "Characterization of mouse epithelial protein lost in neoplasm (EPLIN)
and comparison of mammalian and zebrafish EPLIN.";
RL Gene 262.155-160(2001).
CC -! SUBCELLULAR LOCATION: Cytoplasmic. This cytoskeletal protein co-
localizes with actin stress fibers.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Beta;
CC IsoId=Q9ERG0-1; Sequence=Displayed;
CC Name=Alpha;
CC IsoId=Q9ERG0-2; Sequence=VSP_003118;
CC -! TISSUE SPECIFICITY: Eplin-alpha is highly expressed in embryos
from day 7-11 and in adult spleen and lung. Eplin-beta expression
is highest in adult kidney, testis, lung and liver, intermediate
in heart, brain, spleen, skeletal muscle and low in embryos.
CC -! SIMILARITY: Contains 1 LIM zinc-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
ENBL; AF307844; AAC31147.1;
EMBL; AF307845; AAC31148.1;
HSSP; P32965; ICTL.
GO; MGI:1920992; Eplin.
GO; GO:0015629; C:actin cytoskeleton; IDA.
InterPro; IPR001781; LIM.
Pfam; PF00412; LIM; 1.
ProDom; PD000094; LIM; 1.
SMART; SM00132; LIM; 1.
PROSITE; PS00478; LIM DOMAIN 1; 1.
PROSITE; PS50023; LIM DOMAIN 2; 1.
Cytoskeleton; LIM domain; Metal-binding; Zinc; Alternative splicing.
DOMAIN 386 446 LIM.
VARSP LIC 1 160 Missing (in isoform Alpha).
```

```
FT 216 216 /FTId=VSP_003118.
FT CONFLICT 216 216 S -> N (IN REF. 1; AAC31148).
FT CONFLICT 325 325 A -> T (IN REF. 1; AAC31148).
FT CONFLICT 486 486 P -> S (IN REF. 1; AAC31147).
FT CONFLICT 499 499 V -> A (IN REF. 1; AAC31147).
FT CONFLICT 538 538 S -> G (IN REF. 1; AAC31148).
FT CONFLICT 605 605 T -> A (IN REF. 1; AAC31148).
SQ SEQUENCE 753 AA; 84089 MW; 944692E86DF4A2AF CRC64;

Query Match 76.0%; Score 2983; DB 1; Length 753;
Best Local Similarity 76.3%; Pred. No. 7e-134;
Matches 580; Conservative 60; Mismatches 112; Indels 8; Gaps 3;

QY 1 MESSFPNRRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAAEETNMKKRSNTENL 60
Db 1 MESTFPNRRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAAEANMERKKNPESL 60

QY 61 SQHFRKGTILTVLKKWENPGLCAESHTLSLRNSSTEIRHRDHPHAEVTSHAASAKAQD 120
Db 61 PQHFRGTLSVLKKWENPVAGAEFTDSLPNSSSEGHTADYPPAEVTDKXPAQGVADR 120

QY 121 EEOIHPRSLRSPPEALVQGRYPHIKDGEDLDKDHSTESKMCNCLGESRHEVEKSEISEN 180
Db 121 EHTQPKPRGSRPEAVIQSRIYRSENHSHDFKAQATESQKMCNCLGDSRHEAEKPESEN 180

QY 181 TDSAGKIEKYNVPLNRLKMWFEKGPTQTKILRAQSRGASGRKISSENSYSLDDLEIGPQ 240
Db 181 TTSCKIEKYNVPLNRLKMWFEKGHNQTKSLWTSRNGAGRLSENCSLDDWIGAGH 240

QY 241 LSSSTFDSEKNESRRNLELPRLETSIKDRMAKYQAAVSKOSSSTNYTNELKASGEIKI 300
Db 241 LSSSAFNSEKNESRRNLELPRLETSIKDRMAKYQAAVSKOSSPSYTNELKTS--ESKT 298

QY 301 HWEQKENVPPGPEVCITHOGEKISANENSLAVRSTPAEDSDRSQVSKSEVQVPHKPK 360
Db 299 HWEQKENVPPGPEACSVHQEGSVATENSILVALSVPAEDDTCNSQVSKSEVQVPHKPK 358

QY 361 LSPDRASSLSSESPPKAMKVFQAPARETCVECKTVYPMERLLANQOVHITSRCFSVC 420
Db 359 LSPDARTSSLPESSPKTAKKQAPAKESCVECKTVYPMERLLANQOVHITSRCFSVC 418

QY 421 NNKLSLGTYSALHGRIYCKPHFNLFKSGNYDEGFHGRPHKDLWASKNENEELERPAQ 480
Db 419 NNKLSLGTYSALHGRIYCKPHFNLFKSGNYDEGFHGRPHKDLWASKNENEELERPAQ 478

QY 481 LANARETHSPGVEDAPIAKYCVLAASMEAKASQOEKEDKPAETKRLRIAMPPTLGS 540
Db 479 PPNAGESPHSPGVEDAPIAKYCVLAASMEAKASQOEREDKPAETKRLRIAMPPTLGS 538

QY 541 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLKLRLSSSLKRSRPTTVAASFOST 600
Db 539 SGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLKLRLSSSLKRSRPTTVAASFOST 598

QY 601 SVKSPKTVSPPIRKWMSQSESVGGRVAERKQVENAKASKKNGVNGVKTITWQNKESK 659
Db 599 SVKSPKTVSPPIRKWMSQSESVGGRVAERKQVENAKASKKNGVNGVKTITWQNKESK 653

QY 660 ETGKRSEKHSLENNENLVENGADSDDEDDNSFLKQSQPOEPKSLNWSFVDNTFAEFT 719
Db 654 GEEVPRSKDRSSFELESENFENANGAETDNDHVAQSPLEPAQPGHSGFVDTAAKEF 713

QY 720 TONQKSDQVELWEGEVVKELSVBEQIKRNYDEDEDEE 759
Db 714 TONQKSDQVELWEGEVVKELSVBEQIKRNYDEDEDEE 753

RESULT 3
NFW_HUMAN
ID NFW_HUMAN
AC P07197;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M) (Neurofilament 3).
OS NEF3 OR NEFM OR NFM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87275953; PubMed=3608989;
RA Myers M.W., Lazzarini R.A., Lee V.M.-Y., Schlaepfer W.W., Nelson D.L.;
RT "The human mid-size neurofilament subunit: a repeated protein
RT sequence and the relationship of its gene to the intermediate
RT filament gene family";
RL EMBO J. 6:1617-1626(1987).
RN [2]
RP SEQUENCE OF ONE OF THE 13 RESIDUE REPEATS.
RX MEDLINE=8815820; PubMed=2450354;
RA Lee V.M.-Y., Ovros L. Jr., Carden M.J., Hollosi M., Dietzschold B.,
RA Lazzarini R.A.;
RT "Identification of the major multiphosphorylation site in mammalian
RT neurofilaments";
RL Proc. Natl. Acad. Sci. U.S.A. 85:1998-2002(1988).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Y00067; CAA68276.1; --
CC PIR; A27864; A27864
CC Genew; HGNC:7734; NEF3.
CC MIM; 162250; --
CC GO; GO:0005883; C:neurofilament; TAS.
CC GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
CC InterPro; IPR006821; Filament_head.
CC InterPro; IPR001864; IF.
CC InterPro; IPR002957; Keratin_I.
CC Pfam; PF00038; filament; 1.
CC Pfam; PF04732; filament_head; 1.
CC PRINTS; PRO1248; TYPE1KERATIN.
CC PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW Glycoprotein.
FT INIT_MET 0
FT DOMAIN 1 103 HEAD.
FT DOMAIN 104 411 ROD.
FT DOMAIN 412 915 TAIL.
FT DOMAIN 412 915 TAIL.
FT DOMAIN 104 135 COIL 1A.
FT DOMAIN 136 148 LINKER 1.
FT DOMAIN 149 247 COIL 1B.
FT DOMAIN 248 264 LINKER 12.
FT DOMAIN 265 286 COIL 2A.
FT DOMAIN 287 290 LINKER 2.
FT DOMAIN 291 411 COIL 2B.
FT DOMAIN 613 690 6 X 13 AA TANDEM REPEATS.
FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).

SQ SEQUENCE 915 AA; 102317 MW; 5F5D3DF34C9D9E50 CRC64;
Query Match 6.1%; Score 239; DB 1; Length 915;
Best Local Similarity 22.1%; Pred. No. 0.0002;
Matches 180; Conservative 117; Mismatches 318; Indels 198; Gaps 39;
QY 15 SLRVTALEXSLVNNKSSAIVEIFSYQKAAEETNMKKESNTE----NLQSHFRKGTLT 70
DB 204 ALRKDIEEASLVK-----VELDKKVSQIQDEVAF--LRSNHEEEVADLLAQIASHIT 254
QY 71 VLKKWENPGLG-----ESHTDSLNSSTE---IHRADHPPEAVTSHAASGAK- 117
DB 255 VERKDYLTDTISTALKEIRSQLSHSDQNMHQAEWFKCRKYAKLTEAAEQNKAIRSAKE 314
QY 118 --ADQEQIHPRS----RLRSPPEAL-----VQGRYPHIKDGEDLKDHSSTKKNWENCL 165
DB 315 EIAEYRRQLQSKSIELESVRGTESLERQLSDIEERHNN-----DLSSYQDTTQOLENEL 369
QY 166 GSRHHEVEKSEISENTDASGKIEKYNVPLNRLKNMFEKGEPTQTKILRAQSRASGRKIS 225
DB 370 RGTKEWAR-----HLEYYQDLLN-VKMAIDIEIAAYRKLEEGE-----TRFS 412
QY 226 ENSYSLDLEIGFQLSSSTFDSEKNESRNLRLPEL-----SETSIKRMAYK 274
DB 413 TFAGSI-----TGPLYTHRPDITTSKIQTKEAPKLVQHKFVEIIBETKVEDEKSEM 468
QY 275 QAAVSKOSSSTNTNELKASGGIKIHQMEQKNVP-----PCPEVCIT 318
DB 469 EEAL-----TAITEELAAAMKEEKEAAEPEEPEEBAEVEEVAAKKSPYKATAPEV--K 520
QY 319 HQGEK-----ISAMENSLAVRSTPAEDSDRSQVKSQVQVHPKPLSPDSRASLSLSE 372
DB 521 EEQGEKEEESGQEEEDGAKSDQAEESGSEKESKEGEQEE--GETEAEAGE 577
QY 373 SSPPKAMKFPAPARETCVQKTVPMERLLANQQVFHISCFRCYCNKSLGTVASL 432
DB 578 EAAKEEKKVEEKSEEVATK-----EELVADAKVE-----607
QY 433 HGRIYCKPHNQLFKSGNVDGFGHRPKOLWASKNENEELERPAQLANARETPHSPG 492
DB 608 -----KPEKAKGPVKSPVEEK-GKSP-----VPKSPVEEKGKSPVKSPVEEKGKSP- 654
QY 493 VEDAPITAKVG---VLAASMEAKASSQOEKEDPAETKURIAHPPPTTELGSSSALSEGI 549
DB 655 VPKSPVEEKGKSPVKSPVEEKAQSPVK--SPVEEAKSK-----AEVG-KGEQEEEEE 705
QY 550 KMSKPKWPPPEDEI-SKPEVPEDVDLKLRLRSSLSLKER--SRPFTVAASFQTSVKSPK 606
DB 706 KEVK-EAPKEEKVEKEEKPKQVP---EKKAESPVKEEAFAEVVTTITKSVKVLHEKETK 761
QY 607 TVSPPIRKQMSMSEQSEESVGG--VAERKQVENAKAKK-----NGNVGKTTWQNKESK 659
DB 762 EEGKPLQQ-----EKEKEKAGGEGGSEESGSKGAKSRKEDIYNGEV-----EGK 808
QY 660 GETGKRSKEGHSLEMNENLVENGAD---SDEDDNSFLKQSPQEPKPSLNSWFSVNTFA 716
DB 809 EYVEQETKEGSGREBEKGVYTNGLDLSPADEKKGDKSEKVVVTKVEKITSEGGDGA 868
QY 717 EEFITTON---QKSQD-VELWEGEVVKLSVIE 744
DB 869 TKYITKSVTVTKVBEHETTFEKLVTSTKKVEK 901
RESULT 4
NP14 RAT
ID NP14 RAT STANDARD; PRT; 704 AA.
AC P41777;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body
DE phosphoprotein 1).

Matches 161; Conservative 112; Mismatches 288; Indels 227; Gaps 311.

```

QY 30 KSSAIIVEIFSYYOKAAAEETNMWKKRNTENLSOHFRKGLTIVLKKWENPGLCAE-----84
D 23 RDNLSEVASKFAKATCATQODANASSLLDIYFWLKST-KAPKVKLQSGNPVAKKAKKE 81
QY 85 -SHTDSLNRNSSTIIRHADHP-----AEVTSHAASGA-----116
D 82 TSSDSASEDSSE-EKAQVPTOKAAPAKRASLPQHAGKAAAKAKASESSSESSSEEE 140
QY 117 -----KADBEQIHPRSR-LRSPPEALVOGRYPHIKDGEDLKDHS*ESKKNWENCLGESRH 170
D 141 KDKKKPVQOKAVKPPPK-----KAESSESESDSSEDEAPQTKPKAAA 192
QY 171 EVKSEISENTDASGKIEKYNVPLNKLWMFEKGEPTQTKILRAQSRASGRKISSENSYS 230
D 193 TAAKAPTAKQTKAPK-----PGPPA-----KAQPKAANGKAGSSSS-- 229
QY 231 LDDLEITGPGLSSSTFDSKNEISRRNLELPRISETSIKORM-----AKYQAAVSKOSSST 295
D 230 -----SSSSSSSDSSEKKAAAP-LKKTAPKKVQVAKAPVKVTAAPTQKSSSS 277
QY 286 NYTNELKASGEIKIHMKQKE-----NVPPGPEYCIITHGEKEKISANENSIAVSTPA 339
D 278 E-----DSSSEBEEOKPMKKKAGPYSSVPP-PSVLSK---KSVGAOSP KAAAQTOP 328
QY 340 EDDSRDSQVKSVEQVPHPKPSPDGRASLSSESSPPKAWKFOAPARETCVECKTVP 399
D 329 ADSASD-----SESDSSSEBEKKTPAKTVVSKTPAKPAPVK-----366
QY 400 MERLLANQQVFHISPCRCYCNNKLSLGTYSALHGRYICKPHNQLFKSGKNYDEGFGR 459
D 367 -----KKAESSSDSS-----DSDSSEDEA-----385
QY 460 PHKDLWASKNENEIEIERA-----OLANARETPHSP-GVEDAPIAKVGVLAASMEAKAS 513
D 386 PAKPVSATKSP-----LSKPAVTPKPPAAKAVATPKQAPAGSGQKPKQSKADSSSSESS 441
QY 514 SQEKEKDKAETKYLRI-----AWPPPELSSGSGALBEGIKMSKPKPPPEDEISK 564
D 442 SEEEATKSVTPPKARVATAKAPSLPAKQAPRAGGSDSSSDSSSESEBEKKTTP-----K 496
QY 565 PEVPEDVDLKLKLRSSSLKERSRPFVTAASFQTSVKSPKTVPPPIKGMKSMSEQSEE 524
D 497 PPA-----KKKAAGAAVPKPTPVKKAASSESSSSSDSSSEDESEBEKKK-PSKATPK 546
QY 625 SVGGRVAERKQVENAKASKNGNVGKTTWONKESKG-ETGKRSKEGHSLEMENENLVENG 683
D 547 PQANGKANGVPASONGKAGKESSEEBEDTQNKAAAGTKPGSGKKRKHN-----ETA 597
QY 684 ADSDDNNSFLKQQSPQEPKSLNWSFVNDTFAEEFTTQNKSQDVVELWEGEVVKLSVE 743
D 598 DEEAATPQSKVKVLQTP-----NTFPKSKGKEKRASSPFRVREE---EIEVD 641
QY 744 EQIKRNRY 751
D 642 SRVADNSF 649

```

RESULT 5

ATTRX_MOUSE	STANDARD;	PRT; 2476 AA.
ID	ATTRX_MOUSE	
AC	Q61687;	
DT	15-JUL-1999 (Rel. 38, Created)	
DT	15-JUL-1999 (Rel. 38, Last sequence update)	
DT	15-SEP-2003 (Rel. 42, Last annotation update)	
DE	Transcriptional regulator ATRX (X-linked nuclear protein)	
DE	(Heterochromatin protein 2) (HP1 alpha-interacting protein) (HP1-BP38 protein).	
GN	ATRX OR XNP OR HP1BP2.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	

NCBI_TaxID=10090;

[1]
SEQUENCE FROM N.A.
MEDLINE=98213653; PubMed=9545503;
Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.;
"Comparison of the human and murine ATRX gene identifies highly
conserved, functionally important domains."; *Mamm. Genome* 9:400-403(1998).

[2]
SEQUENCE OF 325-1176 FROM N.A.
MEDLINE=97133299; PubMed=8978696;
le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,
Jeanmougin F., Løsson R., Chambon P.;
"A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic
control of transcription by nuclear receptors."; *EMBO J.* 15:6701-6715(1996).

[3]
SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
HETEROCHROMATIN.
MEDLINE=20040663; PubMed=10570185;
McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
Ackmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
"Localization of a putative transcriptional regulator (ATRX) at
pericentromeric heterochromatin and the short arms of acrocentric
chromosomes."; *Proc. Natl. Acad. Sci. U.S.A.* 96:13983-13988(1999).

-1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
GENE EXPRESSION BY AFFECTING CHROMATIN.

-1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND
PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (By
similarity).

-1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
INTERACTING WITH HP1.

-1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.

-1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.

THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>
OR SEND AN EMAIL TO license@isb-sib.ch)

ENBL; AF026032; AAC08741.1; -
ENBL; X99643; CAA67962.1; -
MGD; MGI:103067; Atrx.
GO; GO:0000228; C:nuclear chromosome; IDA.
InterPro; IPR001410; DEAD.
InterPro; IPR001650; Helicase_C.
InterPro; IPR003330; SNF2_N.
InterPro; IPR001841; Znf_Fing.
Pfam; PF00271; helicase_C; 1.
Pfam; PF00176; SNF2_N; 1.
SMART; SM00487; DEXDC; 1.
DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
Zinc-finger.

ZN FING 219 267 PHD-TYPE.
T NP BIND 1579 1586 ATP (POTENTIAL).
T SITE 1704 1707 DEGH BOX.
T DOMAIN 1704 322 POLY-SER.
T DOMAIN 735 738 POLY-SER.
T DOMAIN 1001 1004 POLY-GLU.
T DOMAIN 1130 1135 POLY-SER.
T DOMAIN 1182 1185 POLY-SER.
T DOMAIN 1238 1245 POLY-ASP.
T DOMAIN 1484 1487 POLY-GLU.
T DOMAIN 1924 1931 POLY-SER.
T DOMAIN 2205 2208 POLY-LYS.
T DOMAIN 2245 2248 POLY-GLU.
T DOMAIN 2403 2408 POLY-GLN.

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M) (Fragment).
GN NEF3 OR NEFM OR NFM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID:9913;
RN [1]
SEQUENCE FROM N.A.
RA Hill W.D., Zhang L., Balin B.J., Sprinkle T.J., Spicer K.,
RA Gearhart D.A.;
RA "The bovine neurofilament M subunit has a novel set of KSP repeats
RT normally restricted to NF-H.";
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RC
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER (BY SIMILARITY).
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF091342; AAC36357.1; -.
CC InterPro; IPR001664; IF.
CC InterPro; IPR002957; Keratin_I.
CC Pfam; PF00038; filament_1.
CC PRINTS; PR01248; TYPEIKERATIN.
CC PROSITE; PS00226; IF; 1.
CC Intermediate filament; Coiled coil; Neurore; Phosphorylation.
CC NON TER 1
CC DOMAIN <1 296 ROD.
CC DOMAIN 297 810 TAIL.
CC DOMAIN <1 20 COIL 1A.
CC DOMAIN 21 33 LINKER 1.
CC DOMAIN 34 132 COIL 1B.
CC DOMAIN 133 149 LINKER 12.
CC DOMAIN 150 171 COIL 2A.
CC DOMAIN 172 175 LINKER 2.
CC DOMAIN 176 296 COIL 2B.
CC DOMAIN 503 582 8 x 13 AA TANDEM REPEATS.
CC SEQUENCE 810 AA; 90799 MW; B8477D85560AC3F6 CRC64;
CC
CC
CC Query Match 5.5%; Score 216; DB 1; Length 810;
CC Best Local Similarity 20.3%; Pred. NO. 0.0021;
CC Matches 170; Conservative 112; Mismatches 320; Indels 234; Gaps 37;
CC
CC 15 SLRVTAKELSLVNKNKSSAIVEIFSQYKAAAEETNWEKKRSNTE---NLSOHFFKGLT 70
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC 89 ALRKDIESSLVK-----VELDKVQSLQDEVAF--LRNHEEVEDADLLAQIQASHIT 139
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC 71 VLKKWENPGLGA-----ESHTSLRNSSTE---IRRHADHPPAEVTSHAASGAK- 117
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC 140 VERDYLKTDISTALKEIRSQLESHSDQNMQAEEWFKCRYAKUTEAAEQNKAIRSAKE 199
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC 118 --AQEEQIHPRS----RLRSPEAL-----VQGRYPHIKGDGLDKHSTSEKMNCL 165
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```
Db 200 EIAEYRQLOSKSIELESVRGTKESLEROLSDIEERHNNH-----DLSSVQDTIQOLENEL 254
QY 166 GESRHEVKEISENTDASGKIEKYVNPVLRNKKMPEKGEPTQTKILRAQSR--SASGRK 223
Db 255 RGTKEWAR-----HLREYQDLN-VKALDILIAAYRKLLEGEETRFRTFAGS 302
QY 224 ISENSYSLDLLEIGPQOLSSSTFDSKNSRNLEPLRL-----SETSIKDRMA 272
Db 303 ITGPLYT---HRQPSIAISSKIQTK-----VEAPKLKVQHGFVEBIEETKVEDEKS 352
QY 273 KYQAAVS--KQSSSTNYTHELKASGGEIKIHKMEQKENY-----PPGEVETITQIE 321
Db 353 EMBEALTAITEELAVGVKEVEEAEERKEEAEAEVVAAKGFVKATAPE--LKSEE 410
QY 322 GEK-----ISANENSLAVRSTPAEDSDSDSVKSEVQVPHPKPLSPDRASSLSSESP 375
Db 411 GEKEEZEGESEEEBEAEKSDQAEQEGGSEKESGESEGESEGESEGESEGESEGESE 470
QY 376 PKAMKXFOAPARETCVECTVYVMBELLANOQVPHISCFRCYCNKLSLGTAYASLHGR 435
Db 471 AKEBKWEKAE-----VAPKEELAAEAKV-----496
QY 436 IYCKPHFNQLFKSGNYDGFGRPHKDLWASKVNEEILERPQALANARETPHSPGVE- 494
Db 497 -----EKPEK-----AKSPVAKSPTTKSPPTAKSPEAKSPEA 527
QY 495 -----DAPIAKVGVLAAAMEAKASSQOEKEDKPAETKRLRTAWPPP-----535
Db 528 KSPTAKSPTAKSPVAK-SPTAKSPEAKSPEAKSPTAKSPTAKSPTAKSPTAKSPTAKS 586
QY 536 -----TELSSGSALBEGIKMSKPPPEDEISKPEVPEDVDLDLKK-----LRRSSSLKE 586
Db 587 KABAGAEKGEQKEVEEKEAKESKPEEKAEEKPKDVP-EKKKABSPVYKASPVKE 645
QY 587 R--SRPFTVAASFQSTSVSKPTVSPPIKKGWSMEQSEESVGRVAERKQVENAKASKK 644
Db 646 EVPAKPKVKVSP-----KEAKEEKPQEK--EKEKEKVEEVGG-----KEEGGLKESRK 692
QY 645 -----NGNVGKTTWQNKSGKETGKSGKSHLEMEENLVENGADSDDDNSFLKQKSP 699
Db 693 EDIAINGEV-----EGKEEETKEKSGSGE-----BEKGVWINGLDVSPGDE--KKGDD 740
QY 700 QEPKSLNWSFVDNTFAE-----BFTTQNKQSD-VELWEGEVVKLSVEE 744
Db 741 KSEKVVVTKWKEKISSEGDDGATKVIKSVTVTKVVEHEETFEKLSTKKVEK 796

RESULT 8
YF16 YEAST STANDARD; PRT: 1233 AA.
AC P43537;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 137.7 kDa protein in UGS1-PAB1 intergenic region.
GN YF016C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -!- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; D50617; BAA09255.1; --
DR PIR; S56271; S56271.
DR SGD; S0001912; YF016C.
KW Hypothetical protein.
SQ SEQUENCE 1233 AA; 137697 MW; CBA7CD2CGF0892F6 CRC64;

Query Match 5.4%; Score 214; DB 1; Length 1233;
Best Local Similarity 20.5%; Pred. No. 0.0042;
Matches 183; Conservative 125; Mismatches 308; Indels 278; Gaps 46;

QY 43 KAABETNMKKRSN-TENLSQHFRKG-----TLTVLKKWENPGLGAESHTDS 89
Db 334 KDVESESLTKGNFNFKENESKHLKAGEKQQTESDRDGISPSVLAKNQKETEIGKEDHVE 393
QY 90 LRNSSTE-----IRHRADHPPAEVTSHAASGAKAD-----QEEQIHPRSRL 130
Db 394 QKQDEKCKRKLSEVNHNN-----MSHNFNAAGSDSIIPPETERETYDDETMGPTKRI 447
QY 131 RSPPEALVQGR-----YPHIKDGEDLKDHSTESK-KMENCLGES-----RHEV---EKS 175
Db 448 SDNEKNLQHGNDISVEVEKEEEEEEEENSTFSVKYKENVTGEOEAARNNEVSGTEEE 507
QY 176 EISENTDASGKIEKYVNPVLRNKKMPEKGEPTQTKILRAQSRGASGRKISSENSYLDLLE 235
Db 508 STSGKEEIMGDEKQS-----EAGE--KSSIIEIG-SANSAKISKDNLVLEDEA 554
QY 236 IGPQLSSST-----FD-----SEKNESRNLELP-----RLSE--T 265
Db 555 EAPTQENKPEWGEIDIPADPRDDVVEIWEAVEENIIPEDLEVAKEQGEQVXLDEPVK 614
QY 266 SIKPRMAYQAAVS-----KQSSSTNYTNEL-----KASGGGIKIHMEQKE 307
Db 615 AMKDDKIAMRGAEISSEDMKKKQEGTAELSNKAKKEVDETAESAEQVEV-----EKS 668
QY 308 NVPPGPPEVC--ITHQEGEKISANENSLAVRSTPAEDDSR--DSOVKSEV-----352
Db 669 KTPESPKVVRCTSGRPEDQINERDPEV---LKEDVRVPDEDVKPEIATTIENSEED 724
QY 353 -----QQPVHPKPLSPDSRASSLSSESPPKAMKKFQAPARETCVQCQTVVPM 401
Db 725 PKSORVQISTEQAETTKQDMGDVGSTTSFKKEEKPKRFEITQEGDKITGD-----775
QY 402 RLLANQVFIHISCFRCYCNKLS-LCTYASLHGRIYCKPHFNQLFKSGNYDEGFGRHP 460
Db 776 ---TNHE--HGEATEAASSENSKASDVGT---AEKYIEPSESSEVKD-----T 814
QY 461 HKDLWASKVNEEILERPQALANARETPHSPGV-----EDAPI-----498
Db 815 BEDAEVENSEKTEFIKVKAELENL-DAPKEAVTAEUNKENEDVEVDTEEDAEVENSEKT 873
QY 499 -----AKVGVLAAAMEAKASSQOEKEDKPAETKRLRIAWPPPTFELSGSSALAEQKMS 552
Db 874 EPIKVKAEGLNLDAPKAEVTAELUNKENEDVEV-----AATSKEDIETKCS 919
QY 553 KPKWPP-----EDEISKPE---VPEDVDLDLKKLRRSSSLK-----ERSRPFTVAAS 596
Db 920 EPABTPLEDGCTCAEAVSKKDAEAVTKEDENMENSKI--AEALKDVTGDOEIDIDINISDE 977
QY 597 FQSTSVKSPKTVSPPIKKGWSMEQSEESVGRVAERKQVENAKASKNGVKTQWQ- 655
Db 978 FQ-----RTVELP-----ELEKQIDKNKGDEKELEVEETEKETSLPDLVVEENI 1022
QY 656 KESKGETGKSGKEGHSLDEM-ENENLVENGADSDDDNSFLKQKSPQEPKSLN---WSSF 710
Db 1023 TEENKVEIKQEEVEVSLQDFNETESISKEAPNNDE--NGFEDQSTRENPKKASADDFIKDI 1080
```

QY 711 VNTFAEFTTQKQSDQVELWEGEVKLSVE-----EQIKENRYDDEDED 757
Db 1081 LDET--NEFLEQKIVDDSEL--NALLQSLDAKSTTQTTEQSKKN--DKPQD 1128

RESULT 9
UN89_CABEL
ID UN89_CABEL STANDARD; PRT; 6632 AA.
AC 001761; Q17362;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]_TaxID=6239;
SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RA MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
assembly, encodes a giant modular protein composed of Ig and signal
transduction domains.";
RL J. Cell Biol. 132:835-848(1996).
RN [2]
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofibrillar
lattice assembly begins with positional cues laid down in the
basement membrane and muscle cell membrane. UNC-89 responds to
these signals, localizes, and then participates in assembling an
M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U33058; AAF00542.1; --
CC EMBL; AF003131; AAB54132.2; --
CC PDB; 1FHO; 20-DEC-00.
CC WormPep; C09D1.1; CE30426.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003598; Ig-C2.
CC InterPro; IPR003006; Ig-MHC.
CC InterPro; IPR001849; PH.
CC InterPro; IPR007850; RCSD.
CC InterPro; IPR000219; RCGD.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00041; fn3; 1.
CC Pfam; PF00047; Ig; 47.
CC Pfam; PF00169; PH; 1.
CC Pfam; PF05177; RCSD; 5.

DR Pfam; PF00621; RhGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SMO0408; IGc2; 23.
DR SMART; SMO0325; RhGEF; 1.
DR SMART; SMO0326; SH3; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS50835; IG-LIKE; 49.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 PH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.

FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFLICT 2137 2137 A -> P (IN REF. 1).
FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).
FT CONFLICT 2258 2258 A -> P (IN REF. 1).
FT CONFLICT 2284 2284 E -> G (IN REF. 1).
FT CONFLICT 2297 2297 M -> I (IN REF. 1).
FT CONFLICT 3531 3531 A -> G (IN REF. 1).
FT CONFLICT 3884 3888 DAGEY -> RRRKI (IN REF. 1).
FT CONFLICT 3929 3929 A -> V (IN REF. 1).
FT CONFLICT 5134 5134 A -> P (IN REF. 1).
FT CONFLICT 5145 5145 T -> S (IN REF. 1).
FT CONFLICT 5185 5185 G -> A (IN REF. 1).
FT CONFLICT 5199 5199 K -> N (IN REF. 1).
FT CONFLICT 5202 5202 L -> F (IN REF. 1).
FT CONFLICT 5213 5213 F -> L (IN REF. 1).
FT CONFLICT 6178 6178 A -> G (IN REF. 1).
FT CONFLICT 6268 6268 K -> E (IN REF. 1).
SQ SEQUENCE 6632 AA; 731665 MW; 262D3EDD62960E89 CRC64;

Query Match 5.4%; Score 213; DB 1; Length 6632;
Best Local Similarity 21.3%; Pred.No.0.034;
Matches 178; Conservative 97; Mismatches 325; Indels 234; Gaps 36;

QY 21 KELSLYNNKSSALVIFSKYQAASETNKKESNTENLSQHRFGTILYVKKWENFG 80
DB 1085 KQSDRVEIRFDGSIKISIKNIKIDAGEIRAVATNSEGDE--TKAKUTVQKKP---- 1138
QY 81 LGAESHTDLSRNSSTEIRHADPHPAEVTSHAASGAKAQEQEIHRPSRLRSPPEALVQ 140
DB 1139 ----APEFDLRPVSLTVKQSE---AVFSHAHGIPLPYYSVNGR-KVRQGEQEARVT 1190
QY 141 RYPHIKDGEDL-----KDHSTESKKMNCILG-----ESRHE---VEKSE 176
DB 1191 RDESTVDGASILTIDATYVSEVNHUTISVVAENTLGAETGAQLIEPKKESVVVEKOD 1250
QY 177 ISENTASGKIEKYNPLNKLKMPKEGP-----TOTKIL----- 212
DB 1251 LS-----SSEVQK-----EIAQQVKEASPEATTTTMTSETLTSTMTTEVTSTVG 1299
QY 213 -----RAQSRAS-----GRKISENSYSDDLEIGPCQLSSSTFDSEKNESSR--- 255
DB 1300 GVTVEPKESSESATTVIGGGSGVTEGSISSVKIEV-VSKTDSQTDVREGTPKRRVSPA 1358
QY 256 NLEPLRLSETSISKDRMAKYQAAVSKOSSSTNYTNELKAS-----GGEIKIHRMEQKENVP 310
DB 1359 EEELPK--EVIDSDRKKKSPDPKXKSEKTEKRPASPTKKTGEVK-----SPKEKSP 1412
QY 311 PGVEVITHOEGEKISANENSIAVRTPADDDSRDQVKSVEQVQVHPKPLSPDRS-ASS 369
DB 1413 ASP-----TKKEKSPAEE--VKSPTKKESKSPSPTKKE-----KSPSPTKKTGDE 1457
QY 370 LSSSSPKKA-----MKAFQAPARETCVE--CQKTVYMEILLANQOVF 410
DB 1458 VKESKSPKSTTKKEKSPKEDVKSPVKKESKSPDATNIVEVSESTIEKTTMTTEM--- 1515
QY 411 HISCFRCSYCNKLSLGTVYASLHGRIVCKPHFNQLFKSGKNGVDEGFGPHRDLWASKNE 470
DB 1516 -----THESEBSRVSUKE-----KTPKEDVKRPSTKDKDKSPKSI 1553
QY 471 NEEL-----ERPAQLANARETPHSPG-----VEDAPIAKVG 502
DB 1554 TEEIKSPVKKESPEKVEKPAASPTTKKESPEKPAASPTKKSSENEVKSPKTKKSP--EKS 1611

QY 503 VLAASMEAKASSQOEKEDKP-AETKKLRIAMPP-----PTELGSSGSALEE----- 547
DB 1612 VVEELKSPKESPKKADDDPKSPKTKKEK---SPEKSATEDVKSPTKKESKPEKVEKPTS 1668
QY 548 -GIMSKPKWPPDEISIKPEVPEDVDLDLKKLRSSSLKERSRPFYVAASQTSVSKPK 606
DB 1669 PTKKESSTKTKTDDEVKSPTKKESKSPQTVPEKPAASPTKKEKSPKSVV-----BEVSKPK 1723
QY 607 TVSP-----PIRXGWSMSEOSEESVGRVARKQVENAKASKKNGVNGKTTWQNK 657
DB 1724 EKSPKABEKSPKSPKTKKESKPEKSAAEVKSPTKKESKPEKSAEKPSPKTKKSSPVKM 1783
QY 658 SKGETGKRSKEGHSLE-----MENENLVENGADSDDEDNSFLKQSQPEP 702
DB 1784 ADDEVKSPTKKESKPEKVEKPAASPTKKEKTPKSA-AEELKSPTKKESKSPSP 1836

RESULT 10
USOL YEAST
ID USOL YEAST STANDARD; PRT; 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INT1 OR YDL058W
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN (2)
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X54378; CAA38253.1; --
DR EMBL; L03188; AAB00143.1; --
DR EMBL; U53668; AAB66659.1; --
DR SGD; S0002216; USOL.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR006955; Usol_p115 C.
DR InterPro; IPR006953; Usol_p115_head.

DR Pfam: PF04871; Usol p115 C; 1.
KW Pfam: PF04869; Usol p115 head; 1.
FT Transpore; Protein Transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216B9FD4818 CRC64;

Query Match 5.4%; Score 211; DB 1; Length 1790;
Best Local Similarity 21.4%; Pred. No. 0.009;
Matches 179; Conservative 128; Mismatches 296; Indels 234; Gaps 38;

QY 7 NRQWTSLSRLVTAKELSLVNNKSAIVEIFSKYKAAEET--NMEKKRSNTENLSQHF 64
DB 1096 NLKNELETLETSSEKALKEVENEER-LAEKIQLEKEATETKQQLNSLRANLESLEKEH 1154
QY 65 RKGTLVLKKNENPGLAG-----SHTSLRNSSTEIRHRDPPPAEVTSHAA 113
DB 1155 E--DLAAQLKYEEQIANKEQVNEISQLNDEITSTQENESIKKNDELEGEV----- 1207
QY 114 SGAKAQEQIIPRSLRPPALVQGRPHDKGDLKHDSKKNMCNLCGESRHEVE 173
DB 1208 KAMKTSSEEQSNLK--KSEIDAL-----NLQIKELKKQNETNEASLLESIKSYE 1254
QY 174 KS--EISENTDASGKTEKYNVPLNRLKMFEXGEPTQTKILAQSRASGRKISNSVL 231
DB 1255 SETVKIKELQDCNFKEK--EVESEDLKASEDKNKKYLELQKES-----EKIKE--EL 1305
QY 232 D----DLEIGPGLSSSTFDSEKNSRRNLPLRSETSIKDR-----MAKYQAAVSKQ 281
DB 1306 DAKTTELKQLEKITNLKAKESSES-----ELSLRKKTSSEKKAEOLEKLNKEIQIK 1361
QY 282 SSTNTYNELKASGGI-----KIKHNE-----QKENVPPGPEVCITHOGEKISAN 328
DB 1362 NOAFERKLLNEGSTIQEYSEKINTLELIRLQNEULKEAKIDNTRSELEKVSLS 1421
QY 329 ENSL-----AVRSTPAE-----DDSRDSQVSEVQVPHKPLSPDRASSUSE 372
DB 1422 NDELLEEKQNTKSLQDEILSYNDKITRDNKLLSIEDKNKDL--ESLKEQLRAAQESK 1479
QY 373 SGPPKAMKFOAPARETCVECKTVVPMERLLANQVPHISCFRCSCNNKLSLGTYSAL 432
DB 1480 AKVEGLKLESESSKEKAELEKSKEMMKKL-----ESTIESNETELKSSMET---- 1527
QY 433 HGRIVCKPHFNQIFKSKNGNYDEFGHRPHK--DLWASKNENE-EILERPAQLANARETPH 489
DB 1528 ----IRKSDEKLEQSKGAEDIKNLQHEKSDLSIRINESKDIELKSKL----- 1574
QY 490 SPQVEDAPIAKGVVLAAASWEAKASQOE--KEDKPAETKULRIAWPPPTLGGSSGALEE 547
DB 1575 -----RIEAKSGSELETQVQELNNAQEKIRI-----NAEEN 1605
QY 548 GIKMSKPKWPEDEISKPEVPDVLDLK-----KLRRSSSLKE--RSRPFVVAASFQTS 601
DB 1606 TVLKSK-----LEDIERELKQKAEIKSNQOEKELTSLRKELEQELDSTQ 1651
QY 602 VKSPKTVSPPIRKGMWSMQSESVGGRVAERK--OVENAKASKK-----NGNVGK-T 651
DB 1652 QRAQK-----SEEE--RAEVKRFQVEXSQLEKAMLETKNYKLVNKEQ 1694
QY 652 TWQWKE-----SKGTGKRSKEGHSLEMENENLVENGADSDDDNSFLKQSQPOEP 702

DB 1695 AKRDEDTVKKTTDSQROEIEKLAKELDNLKAENSKLKEANEDRSEIDDLML----- 1746
QY 703 KSLNWSSEVDNTFAEEFTTQNKQSQVELWEVGEVVKLSVEEQIKRRRYDEDEDEE 759
DB 1747 -----LVTDLDEKNAKYR-----SKLKDLGVE--ISSDEEDDEDEDEE 1782

RESULT 11
MYH9 RAT
ID MYH9 RAT STANDARD; PRT; 1961 AA.
AC G62812;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain, type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
OS MYH9.
GN Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.; Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS, CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND CAPPING.
CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
DR EMBL: U31463; AAA74950.1; --
DR HSSP: P10587; 1BR2.
DR InterPro: IPR000048; IQ region.
DR InterPro: IPR001609; myosin_head.
DR InterPro: IPR004009; myosin_N.
DR InterPro: IPR002928; myosin_tail.
DR Pfam: PF00612; IQ; 1.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF02736; myosin_N; 1.
DR Pfam: PF01576; myosin_tail; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS00096; IQ; 1.
DR PROSITE: PS00096; IQ; 1.
KW Myosin; ATP-binding; G-actin-binding; Actin-binding; Coiled coil; Alkylation; Multigene family.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1961 AA; 226336 MW; 993876D9681FB19E CRC64;

Query Match 5.2%; Score 206; DB 1; Length 1961;
Best Local Similarity 19.5%; Pred. No. 0.017;

Matches 162; Conservative 130; Mismatches 305; Indels 234; Gaps 33;

QY 14 LSLRVAK--ELSLVKNKSSAIVEIFSKYQKAAEETNMKKRSNTENLSO----- 62

Ds 902 LRLRLTAKKQLEBICHLEARVEEBEERCQYLQAE--KKQVQNIQLEBEQLEESAR 959

QY 63 ---HFRKGTTLVKKKKNPGLGSHSTDSLRNSTRIRHRAHPAETVTSAAAGAKAD 119

Ds 960 QKQLEKVTTEAKLKLBEDQIIMEDQCNKLAKKLLERDV---AEFTDL-----ME 1010

QY 120 QEOIHPRSRLSPPEALVQGRYPHDKGDELKDHSTESKKVNCGLSGSHEVEKSEISE 179

Ds 1011 EEBKSKSLAKLVKHEAMV-----TDLERLAREEKQROELEKTRKLE---GD 1056

QY 180 NTDASGKIEKYNVPLNRLM-MFKGEGTQTKILRAQSRASGR-----KISEN 227

Ds 1057 STDLSQDAELQAOIAELKMLAKKEELQALARVEEAAQKMAWKKIRELETQISEL 1116

QY 228 SYSL-----DLEITGPGOLS---SSTFSEKNEGRNLELPRISSET- 265

Ds 1117 QEDLESEPRACRNKAEKQKRDJGEELEAKTLEDTLDTAAQOELRSKREQEVSLKXTL 1176

QY 266 -----SIKRMVAKYQAAVSKSSSTNYTNELKASGGEIKHKVEQKENVPPGEVC 316

Ds 1177 EDEAKTHEAQIOEMRQKHSQAVEELASQLEOTKVKATLEKAK--QTLENERGELANEVKA 1235

QY 317 ITHOGE-----KISANENSLAVRSTPAEDSDSDQVSKVEQVQVHPKPLSPDSRASSL 370

Ds 1236 LQKQGDSEHKRKKVEAQLOLQVFKSEGE-----RVTELADKVKSLQVELDSVTGLL 1289

QY 371 ---SSSPKPKAMKTKQAPARETCVQCQTKVPMERLLANQVVFHISCFRCYCNKLSLGT 428

Ds 1290 NQSDSKSSKLTQDFA-----LESQLODTQELLOEEN-----RQKLSLST 1329

QY 429 YASLHGRITCKPHENQOLFCKSNVDGFGHPRKDLWASKNENEIILERPAAQLANARETP 488

Ds 1330 -----KUKQMEDEKNSPRE-----OLEEEEEAKRNLEKQIATLHAQVTD 1369

QY 489 HSPGVEDAPIAKVGLAASMEAKASSOEKEDKPAETKRLIAWPPPTLSSGSSALEEG 548

Ds 1370 MKKXNEDG-----VCLTEAEAKRLQKLE-----GLSQRLLEK 1405

QY 549 I-----KMSKPKWPEDEISKEPEVDVLDLKLKRRS--SSLKERSRPFTVAASFQSTSVK 603

Ds 1406 VAAVDKLEKTKTRIQEEL-----DDLVLVDHQVQSVNLEKKQKQKFDQLLA----- 1452

QY 604 SPKTVSPPIRGWMSQSESVSGRVAERKQVENAKSKNG---NVGKTTWQNKESKG 660

Ds 1453 -----EKTISAKYAEERDRAEAFAAREKETKALSLARALEEAMEQKA 1494

QY 661 BTGKRSKEGHSLEMENENLVNGADSD-----EDDNSFLKQSQSPQPKSLNWSFVDNT 714

Ds 1495 ELERLNQO---FRTEMEDLSSKDDVGVKSVHELEKSNRALEQQ----- 1534

QY 715 FAEFTTQNKQSQDVLEWGEVVK---ELSYE---EQIKRNVYDEDEDEE 759

Ds 1535 -VEEMKTOLEBED-ELQATEDAKLRLEVNLOAKMAQOFERDLQGRDQSE 1583

RESULT 12

NFM_MOUSE

AC P08553; 061951; STANDARD; PRT; 848 AA.

DT 01-AUG-1998 (Rel. 08, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Neurofilament triplet M protein (160 kDa neurofilament protein)

DE (Neurofilament medium polypeptide) (NF-M)

GN NEF3 OR NEFM OR NFM

OS Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxId=10090;

RN [1] SEQUENCE FROM N.A.

RP MEDLINE=87246694; PubMed=3036526;

RA Levy E., Liem R.K.H., D'Eustachio P., Cowan N.J.;

RT "Structure and evolutionary origin of the gene encoding mouse NF-M,

RL Bur. J. Biochem. 166:71-77(1987).

RN [2]

RP SEQUENCE OF 322-540 FROM N.A.

RX MEDLINE=87158637; PubMed=3103856;

RA Julien J.-P., Meyer D., Flavell D., Hurst J., Grosfeld F.;

RT "Cloning and developmental expression of the murine neurofilament

RL gene family."

RX Brain Res. 387:243-250(1986).

CC -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,

CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.

CC -I- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPETIDE K-S-P, NFM IS

CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS

CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF

CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE

CC OF AXONAL CALIBER.

CC -I- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING

CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE

CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND

CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.

CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X05640; CAA29127.1; -

DR EMBL; M20481; AAA39815.1; -

DR PIR; B43772; B43772

DR PIR; S00030; S00030.

DR MG; MG1:97314; Ne13.

DR InterPro; IPR006821; Filament_head.

DR InterPro; IPR001664; IF.

DR Pfam; PF00038; filament; 1.

DR Pfam; PF04732; filament_head; 1.

DR PROSITE; PS00226; IF; 1.

KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;

KW Glycoprotein.

FT INIT MET 0 0

FT DOMAIN 1 102 HEAD.

FT ROD. 103 410

FT DOMAIN 411 848 TAIL.

FT DOMAIN 103 134 COIL 1A.

FT DOMAIN 135 147 LINKER 1.

FT DOMAIN 148 246 COIL 1B.

FT DOMAIN 247 263 LINKER 12.

FT DOMAIN 264 285 COIL 2A.

FT DOMAIN 286 289 LINKER 2.

FT DOMAIN 290 410 COIL 2B.

FT CARBOHYD 47 47 O-LINKED (GLCNAC) (BY SIMILARITY).

FT CARBOHYD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).

FT CONFLICT 432 432 S -> F (IN REF. 2).

FT CONFLICT 539 540 QA -> RR (IN REF. 2).

SQ SEQUENCE 848 AA; 95910 MW; 5F251F274D0F13B6 CRC64;

Query Match 5.2%; Score 205; DB 1; Length 848;

Best Local Similarity 19.5%; Pred. No. 0.0072;

Matches 160; Conservative 116; Mismatches 287; Indels 256; Gaps 35;

QY 9 ROWTSLSLRVTKAKLSLVNKNKSSAIVEIFSKYQKAAEETNMKKRSNTE--NLSQHFPRK 66

Ds 194 RQDTEAARALRAUKDI-----EESMVKVELDKVQSLQDEAVFLRNHEEVEADLLAQICA 249

QY 67 GTLTIVKXKKNWPNGLGA-----ESHTSLRNSSTE---IRHADHPPAEVTSHAAS 114

Db 250 SHITVERKDYLTDTISTALKEIRSOLECHSDQNHQAEWFKRYAKLTAAEQNKRAIR 309
Qy 115 GAK---ADQEQIHPRS-----RLRSPEAL-----VQGRYPHIKQGEDLKHSTSEKXM 161
Db 310 SAKETIAYEROLQSKSIELESVRGTKEISLERQISOIERHNH-----DLSSYQDTIQOL 364
Qy 162 ENCLGESRHEVSEISENTDASGKIEKYNVPLNRLKMKMEKEGPTQTKILRAQSSASG 221
Db 365 ENEURGTQWEMAR-----HUREYQDLN-VQMALDIBIAAYRKLLEGE-----407
Qy 222 RKISENSYSLDLEIGP-----GQLSSSTFDSEKNESSRNLELPL-----SETSI 267
Db 408 TRSTFSGSI---TGPLYTHRQPSYTISSKIOKTK---VEAPKLKVQHKEVEEIIETKV 461
Qy 268 KQWAKYQAAVSKQSSSTNTYNELKASGGIKIHKMEQKENVP-----PGPEVCITHQ 320
Db 462 EDEKSEMEETL-----TATAELAAASAKBEK-EAAEKEEPEAEKSPVKSPEAKEEBE 514
Qy 321 EGK-----ISANENSLAVRSTPAEDDSRDQVSEVQOPVHPKPLSPDPSRASSLSSES 373
Db 515 EGKKEEEOEKEEBEEDGVKSDQAEEGSEKESKEDGQEBE-EGETAEAGEGEE 573
Qy 374 SPPKAMKKFOAPARETCVEQKTYVPMERLLANQOVFHISCFRCSCYCNKKLSLGTVASLH 433
Db 574 AEAKEEKIEGKVEEVAVK-----592
Qy 434 GRIYKPHFNQLPKSGNYDEGFGHRPHKDLWASKNEEI-LERPAQANARETSPG 492
Db 593 -----BEIKVEKEKAKS---PMPKSPV 612
Qy 493 VEDAPAKVGLAASWEAKASSQOEKEDKPAETKURIAMPPPTTELGGSSGSALEGIKMS 552
Db 613 EYKVP-----KPEAKAGKEQKEEKEVEKEVTKESPE-----648
Qy 553 KPWPEDEISKPEVEDVDLLKLRSSSLKERS--RPFTVAASFQTSYKSPKTVSP 610
Db 649 -----EKVEKEKPKDV---ADKKAESPVEKAEVETITKSVKVSLEKOTKEEP 699
Qy 611 PIRKGSMSQSESVGRVAERKQVENAKASK-----NGNVGKTTWQNKESKGTGR 665
Db 700 QPOB--KVKEKAEEGGSE--EGSDRSQESKEDIANGEV-----EGKEEEOE 747
Qy 666 SKGHSLEMNENLVENGADSDDDNSFLKQSQPQPKSLNWSFVDNTFAEFTTQOK 725
Db 748 TOEKGSGREBEKGVINGLD-----VSPAEEKKGEDSS-----DDKVVVTK 789
Qy 726 SQDVELWEGE-----VVKELSVQEKIRNYDEDEDEE 759
Db 790 VEXITSEGGDGATKYITKSVTVTKRVEEH-----EETFE 824

RESULT 13
Y67_YEAST STANDARD; PRT; 1658 AA.
AC Q03661; Q04988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 187.1 kDa protein in GUAI-ERG8 intergenic region.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Fukuyota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odeh C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.,

RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.",
RL Nature 387:90-93 (1997).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49809; CA489934.1; -
DR EMBL; Z49939; CA490190.1; -
DR PIR; S55101; S55101.
DR SGD; S0004832; ESC1.
DR GO; GO:0005634; C:nucleus; IDA
DR GO; GO:0006348; P:chromatin silencing at telomere; IMP.
KW Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;
Query Match 5.1%; Score 201; DB 1; Length 1658;
Best Local Similarity 18.2%; Pred No. 0.024;
Matches 173; Conservative 146; Mismatches 312; Indels 320; Gaps 37;
Qy 24 SLVKNKSSAIVEIPSKYKAAETNMKRSN-----TENLSQHFRKGTTLVL 72
Db 94 SLINRGKS--LITLLEK-EDALFERSLEERQRFQHLHSLMKNKYTGNSKSHQR---LIDL 147
Qy 73 KKKWENPGLGAEHTDLSRNSSTEIRHRA--DHPPAEVTSAAAGAKADQEQIHPRL 130
Db 148 RK-----SQYGTDTSFQNNDEIPLDSFISSPLDPAEDSSNIDSKDDELEGKQSL 199
Qy 131 -----RSPPEALV-----QGRYPHIKGE-----DLK 152
Db 200 IKDFLENDEVELSEKNSQGSPPSIMLSDEYAEAGALQDVNSDEYAEEGQVPERK 259
Qy 153 DHSTESKKNVCLG-----ESRHEVEKSEISENTDASGK 186
Db 260 NIGQEQANVENATQISSSDSSEGGQYSEGVMELEDDIDVESDAEKDESQGAETHEHVD 319
Qy 187 TEKNVPL--NRLKMMFEKGEPTQTKI-----LRAQSR 217
Db 320 FSKYMQPRTDKTPIVIEKYESDEHKVHQVSEDGAPFGSVNTSVDDSEDESEQAESY 379
Qy 218 SASGRKI--SENSYSLDLEIGFQGLSSSTFDSEKNESSRNLELPLRSLTSIKDRVAKYQA 276
Db 380 SANAENVVHHNEHLEDDKEL-----IEDIESSSESQSAQESQGSDDDFEYKM-KUEK 432
Qy 277 AVSKQSSSTNTYNELKASGGIKIHKMEOKEN--VPPGPEVCIT-----HQEKEKISAN 328
Db 433 STSEETENTSESROGFAKDAYTKNKVQEQQENDEEPEKDDIIRSLDKNFHGNKNSKSEYS 492
Qy 329 ENSLAVRSTPAEDDSRDQVSEVQOPVHPKPLSPDPSRASSLSSESP-----PKAMKK 381
Db 493 ENVLNETDPAIVE-RENOIND-----VEGVDVTGKSVESDLHHSPLNLYDLAARMLQ 546
Qy 382 FOAPARETCVEQKTYVPMERLLANQOVFHISCFRCSCYCNKKLSLGTVASLHRIYCKPH 441
Db 547 FQGRNSNCPQKEQV-----SESYLCHSNGSNLSGR-----578
Qy 442 PNQLFKSGNYDEGFGHRPHKDLWASKNEEIEILERPAQANARETSPHSGVEDAPIAKV 501
Db 579 -----SLDESEEQIPLKDFGTGENNN-----LKTDR-----604
Qy 502 GVLAASMEAKASSQOEKEDKPAETKURIAMPPPTTELGGSSGSALEGIKMS-----552
Db 605 GDLSSSEIVEIKVSEKK-LDGSTEKELVPLSTDTTINNSLGNEDSIYSLDDADAISE 663
Qy 553 -----XPKM-----PPDEISKPEVPE---DVIDDLK 576
Db 664 NLTDVPLWEIKTTPKYEVVISESVYSSTVEDNTVAMPPOVEYTPPMNDFNSLNDYYE 723

QY 577 KLRSSSLKERSRPTVAASFQSTSVKSPKTVSPPIRGWMSMSQSESVGRVAERKQV 636
Db 724 K--RHLLKS-----TLAALAPATKKDAEFVEAGVTKSCLTSTSGHTNIPHTSKETQV 776
QY 637 ENAKASKKNGVGTITWONKESKGETGRSKE-----GHSLEMEH----- 676
Db 777 SLDLSTEN-----VTFENENTGDKNKQSKNFQGVANSTDKSTEDTNDKYFSAINYTN 831
QY 677 -----ENLVENGADSD-----DMSFLKQSQPQPKSLNWSFVDNT 714
Db 832 VTGSSCCDIETASNVENLRYCEKONMAEMSSGDECVKQNDGKTKQISFSTDSPDN 891
QY 715 FAE-----EFTQKQSDVLMESCE-VVKELSVEEQIKRNYVDEDEDE 759
Db 892 FOESNDNTEFSSTKYKVRNSLDEDESUKKELTKAEVVVK-----LDESESD 939
RESULT 14
ID_MYS2 DICDI STANDARD; PRT; 2116 AA.
AC P08759;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MHCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A. PubMed=3540939;
RX MEDLINE=87092266; PubMed=3540939;
RA Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from
Dictyostelium discoideum";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN=AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippller J.,
RA Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a
phosphorylatable heavy chain fragment of Dictyostelium myosin II";
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
Dictyostelium myosin heavy chain";
RL FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutch K., Holden H.M.,
RA Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
discoideum complexed with MgADP.BeFx and MgADP.AlF4";
RL Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
truncated head of Dictyostelium discoideum myosin to 2.7-A
resolution";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the

Dictyostelium discoideum myosin motor domain to 1.9-A resolution."; Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452590; PubMed=9305951;
RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammaS, and MgAMPgammaP complexes of the Dictyostelium discoideum myosin motor domain."; Biochemistry 36:11619-11628(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148; Bagshaw C.R., Rayment I.; Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization of Mg.2(3')-O-(N-methylanthraniloyl) nucleotides bound to the Dictyostelium discoideum myosin motor domain."; J. Mol. Biol. 274:394-407(1997).
RN [9]
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL CORTEX.
CC -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LAM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1 POSITION (688).
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M14628; AAA33227.1; ..
CC PIR; A26655; A26655.
CC PDB; 1MNA; 03-DEC-97.
CC PDB; 1MND; 17-AUG-96.
CC PDB; 1MNG; 03-DEC-97.
CC PDB; 1MNN; 03-DEC-97.
CC PDB; 1MND; 17-AUG-96.
CC PDB; 1MNE; 23-DEC-96.
CC PDB; 1VOM; 23-DEC-96.
CC PDB; 1LVK; 28-JAN-98.
CC PDB; 1DOX; 20-DEC-00.
CC PDB; 1DOY; 20-DEC-00.
CC PDB; 1DOZ; 20-DEC-00.
CC PDB; 1DIA; 20-DEC-00.
CC PDB; 1DIB; 20-DEC-00.
CC PDB; 1DIC; 20-DEC-00.
CC PDB; 1PMV; 20-DEC-00.
CC PDB; 1PMW; 20-DEC-00.
CC PDB; 1G8X; 17-JAN-01.
CC PDB; 1JWY; 07-NOV-01.
CC PDB; 1XZ2; 07-NOV-01.
CC DictyDB; DD01008; mhca.
CC InterPro; IPR000048; IQ region.
CC InterPro; IPR001609; myosin head.
CC InterPro; IPR004009; Myosin_N.

DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT DOMAIN 1 761
FT DOMAIN 762 791
FT DOMAIN 817 2116
FT NP_BIND 179 186
FT DOMAIN 638 660
FT DOMAIN 738 752
FT MOD_RES 130 130
FT MOD_RES 130 130
FT MOD_RES 678 678
FT MOD_RES 1823 1823
FT MOD_RES 1833 1833
FT MOD_RES 2029 2029
FT TURN 3 5
FT TURN 7 8
FT HELIX 10 15
FT HELIX 25 28
FT STRAND 34 37
FT TURN 41 42
FT STRAND 48 55
FT STRAND 59 63
FT STRAND 69 73
FT HELIX 74 76
FT STRAND 78 79
FT HELIX 83 85
FT TURN 86 87
FT STRAND 90 90
FT HELIX 91 93
FT HELIX 99 110
FT TURN 111 113
FT STRAND 116 119
FT TURN 120 121
FT STRAND 122 126
FT HELIX 137 142
FT TURN 143 145
FT HELIX 148 150
FT HELIX 155 169
FT TURN 170 170
FT STRAND 173 179
FT TURN 181 182
FT HELIX 185 200
FT HELIX 210 226
FT STRAND 227 228
FT TURN 234 235
FT STRAND 236 237
FT STRAND 240 247
FT TURN 249 250
FT STRAND 253 261
FT HELIX 265 268
FT TURN 269 269
FT TURN 273 274
FT STRAND 278 278
FT TURN 279 287
FT TURN 290 296
FT TURN 297 297
FT HELIX 301 303
FT TURN 305 307
FT TURN 316 317
FT HELIX 320 334
FT TURN 335 335
FT HELIX 338 355
FT TURN 356 356
FT STRAND 360 360
FT STRAND 368 368
FT STRAND 373 382

Query Match 5.1%; Score 201; DB 1; Length 2116;
Best Local Similarity 19.9%; Pred. No. 0.032;
Matches 165; Conservative 136; Mismatches 309; Indels 218; Gaps 33;

QY 14 LSLRVTAKELSLVNKNKSSAIVEIFESKY-----QKAAEETNMEKKRSNTENLSOHFRKGT 68
DB 821 LKRRPFEXEI-----KEKEREILELKSNTLDTSTQDKLEKSLKDTESNVLDLQRLQ-KAE 875
QY 62 LTVLKKNWENPGLGAESHTSLRNSSTEIRHADHPPAEV-----TSHAASGA 116
DB 876 KETLKAMYDS-----KDALEAQKREIRVEDMESELDKLEALNQLNQKRSVEEK 927
QY 117 KADQEEQIHPRSLRSPPEALVQGRYPHIKQEDLKHSTES-KWENCLGESRHEVEK- 174
DB 928 VRLDEELQEEQLKNTLEKLYKKEEELMKRYNDQSDTISRLEKIKDELQKEVEEL 987
QY 175 -SEISENTDASKIKYNV-----PLNRLKMFKEGPTQTILRAQ----- 215
DB 988 TESFSESKDKGVLEKTRVRLQSELDLDTVLRLDSETDKSLLKQKKLEELKOVQAL 1047
QY 216 -----SRASGRKISENSYSLDDLEIGPQQLSSSTFDEKKNESRRNLPLRSETSI 267
DB 1048 AAETAAKLAQEAANKL-QGEYT-----ELNEKFNSEVTARNVSKKTKLESQLVAVNE 1102
QY 268 KDRMAKYOAAVSKOSSSTN-----YTNELKASGGEEK-IHKMEQKENVPPGPEVCITHQE 321
DB 1103 LDEKKNRDALEKKKKALDAMLEEMKDQLESTGGCKSLYDLKVK-----QE 1149
QY 322 GEKISANENSLAVRSTPAEDDSRDSQVKSVEQVPHPKPLSPDRASSLSSESSPKAMVK 381
DB 1150 SDMEALRNQISELOSTIAKLEKIKSTLEGEVAR-----LQGELEAEQLAKSNVEKQKK 1203
QY 382 FQAP-----ARETCVE--CQXTVYPMERLLANQOVFHISCFRCSYCNKNKLSLGTVAS 431
DB 1204 VELDLKSAQLAEETAAKQALDKKKLQELSEVQT-----OLSEANNK----- 1249
QY 432 LHGRYICKPHFNQLFKSKGNVDEGHRPHKDLWASKNENEEILERPAQLANARETSPHSP 491
DB 1250 -----NVNSDS-----TNKHLSTFNNLKLEAEQKAKQALEKKR-- 1285
QY 492 QVEDAPIAKGVLAASWEAK-ASSQQEKEDKPAETKRLRIAMPPTTIGSSGSALEEGIK 550
DB 1286 -----LGESELKHVNEQLEEEKKQKSNEKR-----KVDLEKEVSELKQIE 1328
QY 551 MSKPKWPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPTVAASFQSTSVKSPKTVSP 610
DB 1329 -----EEVASKKAVTE-----AKNKKSESELDKIKROYADVVSRRDKSVEQLKTLQ- 1373
QY 611 PIRKGSWMSQSEESVGRVAERKQVENAKSKNG-----NVGKTTWQNK 656
DB 1374 --AKNEELRNTAEBAEG-----QLDRASRKKKAEFDLEEA VKNLEETAKVKKAEKA 1424
QY 657 ESKGETGKRSKEGH-----SLEWENENLVENGA---DSDEDDNSFLKQSQSPQ 700
DB 1425 MKKAETDYRSTKSELDDAKNVSSQYQIKLNEELSELSESVLEEADERCNSAIKAKKTA 1484
QY 701 EPKSLNWSSFVD---NTFAEFTTQNKQSDVELWEGEVVKLSVERQ 745
DB 1485 ESALSLKDBIDAANNAKAK---AERKSKELEVRVAEL--EESLEDK 1526

RESULT 15

MYH9 HUMAN
ID MYH9 HUMAN STANDARD; PRT; 1960 AA.
AC P35579; O60805;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON [1] NCBI_TaxID=9606;
 RN RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter M.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kersey J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghy-Mohammadi M., Matthews L.H., McCann O.T.,
 RA McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Shira H.K., Skuce C.D., Smalley S., Smith M.B.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Miroshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
 RA Korfi I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
 RA Emanuel B.S., Shaikh T., Kuranashi H., Saitta S., Budar M.L.,
 RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
 RA Kim U.J., Shizuya H., Simon M.I., Dunanski J.P., Peytard M., Kedra D.,
 RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
 RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
 RA Tiahun Y., Wright H.;
 RT "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 RN [2]
 RP SEQUENCE OF 1-1337 FROM N.A.
 RX MEDLINE=92003925; PubMed=1312569;
 RA Tothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
 RA Arnaut M.A., Clayton L.K., Tenen D.G.;
 RT "Cellular myosin heavy chain in human leukocytes: isolation of 5'
 RT cDNA clones, characterization of the protein, chromosomal
 RT localization, and upregulation during myeloid differentiation.";
 RL Blood 78:1826-1833(1991).
 RN [3]
 RP SEQUENCE OF 1-715 FROM N.A.
 RX MEDLINE=91316803; PubMed=1860190;
 RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
 RA Ghula D., Adelstein R.S., Weir L.;
 RT "Human nonmuscle myosin heavy chains are encoded by two genes located
 RT on different chromosomes.";
 RL Circ. Res. 69:530-539(1991).
 RN [4]
 RP SEQUENCE OF 714-1960 FROM N.A.
 RX MEDLINE=90138958; PubMed=1967836;
 RA Sasz C.G., Myers J.C., Shows T.B., Leinwand L.A.;
 RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity
 RT through alternative polyadenylation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).

[5]
 RN RP VARIANT DFNA17 HIS-705.
 RX MEDLINE=20489856; PubMed=11023810;
 RA Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
 RA Mhatre A.N.;
 RT "Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in
 RT nonmuscle myosin MYH9.";
 RL Am. J. Hum. Genet. 67:1121-1128(2000).
 RN [6]
 RP VARIANTS MHA/FTNS/SBS LYS-93; CYS-702; CYS-1165; HIS-1424 AND
 RX LYS-1841.
 MEDLINE=20428192; PubMed=1097359;
 RA Seri M., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
 RA Ghiglieri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apolito M.,
 RA Iolascon A., Zellate L.L., Savoia A., Balduini C.L., Noris P.,
 RA Magrini U., Belletti S., Heath K.E., Babcock M., Gluckman M.J.,
 RA Aliprandis E., Bizziaro N., Desnick R.J., Martignetti J.A.;
 RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
 RT Sebastian syndromes.";
 RL Nat. Genet. 26:103-105(2000).
 RN [7]
 RP VARIANTS MHA ILE-1155 AND LYS-1841.
 RX MEDLINE=20428193; PubMed=10973260;
 RA Kelley M.J., Jawien W., Ortel T.L., Koczak J.F.;
 RT "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in
 RT May-Hegglin anomaly.";
 RL Nat. Genet. 26:106-108(2000).
 CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
 CC CAPPING.
 CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
 CC CHAIN SUBUNITS (MHC) 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
 CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGGLIN ANOMALY
 CC (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF FECHTNER SYNDROME
 CC (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
 CC WITH ADDITIONAL ALPORT-LIKE CLINICAL FEATURES OF SENSORINEURAL
 CC DEAFNESS, CATARACTS AND NEPHRITIS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME
 CC (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.
 CC -!- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
 CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA17) WHICH IS
 CC CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND
 CC COCHLEOSACULAR DEGENERATION.
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 282215; CAB05105.1;
 CC EMBL; M81105; AAA59888.1;
 CC EMBL; M69180; AAA61765.1;
 CC EMBL; M31033; AAA36349.1;
 CC PIR; A61231; A61231.
 CC HSSP; P10587; 1BR2.
 CC Genew; HGNC:7579; MYH9.
 CC MIM; 160775;
 CC MIM; 153640;
 CC MIM; 155100;
 CC MIM; 603622;

DR MIM; 605249; ..
 DR GO; GO:0005960; C:non-muscle myosin; TAS.
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR001609; myosin_head.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS00396; IQ; 1.
 KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
 KW Coiled coil; Alkylation; Multigene family; Disease mutation;
 KW Deafness.
 FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
 FT DOMAIN 779 808 IQ.
 FT DOMAIN 837 1926 COILED COIL (POTENTIAL).
 FT NP_BIND 174 181 ATP (POTENTIAL).
 FT DOMAIN 654 676 ACTIN-BINDING.
 FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
 FT VARIANT 93 93 N -> K (IN MHA).
 FT VARIANT 702 702 /FTID=VAR 010791.
 FT VARIANT 705 705 /FTID=VAR 010792.
 FT VARIANT 1155 1155 R -> H (IN DFNA17).
 FT VARIANT 1165 1165 T -> I (IN MHA).
 FT VARIANT 1424 1424 /FTID=VAR 010794.
 FT VARIANT 1424 1424 /FTID=VAR 010795.
 FT VARIANT 1841 1841 D -> H (IN FTNS).
 FT VARIANT 1841 1841 /FTID=VAR 010796.
 FT CONFLICT 53 55 EAI -> RGH (IN REF. 3).
 FT CONFLICT 660 660 T -> S (IN REF. 3).
 FT CONFLICT 869 869 T -> M (IN REF. 4).
 FT CONFLICT 931 931 C -> Y (IN REF. 4).
 FT CONFLICT 1240 1241 KG -> GR (IN REF. 4).

Query Match

Best Local Similarity 5.1%; Score 200.5; DB 1; Length 1960;
 Matches 167; Conservative 151; Mismatches 321; Indels 269; Gaps 34;

QY 14 LSLRVTA--ELSLVNNKSSAIVIFSKYQAAEETNMEKRSNTENLSQ----- 62
 DB 902 LAARUTAKQLEEEICHDLAEARVEEERCOHLQAE--KKKQOQNIQLEEQLEESAR 959
 QY 63 ---HFRKGLTVLKKWENPGIAGSHSDLSLNSSTEIRHRADHPPAEVTSHAASAKAD 119
 DB 960 QKLQLEKVTTEAKLKKLEBEQIILEDQNKLAKEKKLLEDRI---AEFTNL-----TE 1010
 QY 120 QEEQIHPRSLRSPALVQGRYPHIKQGEDIKDHSKSKMENCLGESRHEVEKSEISE 179
 DB 1011 EEEKSKSLAKLKNKHEAMT-----TDLEERLRREKQORQELEKTRKLE-----GD 1056
 QY 180 NTDASGKIEKYNVPLNRLKM-MFEKGEPTQTKILRAQSRSAG-----PKISNSYSLDDL 234
 DB 1057 STDLSQTAELQAQTAELKQAKLEEBELQALARVEEAAQKNMALKKIRLEESQISEL 1116
 QY 235 --EIGFGQLSSSTFFSEKNESNRNLELPLRSETSIKDRMAKYQAQVSKSSSTNY---TN 289
 DB 1117 QEDLESERASRNKAQKQEDLGEELAEALKTELEDLTDSTAAQQLRSRQGEVNLKCTL 1176
 QY 290 ELKASGGEIKIHMFOKE-----NVPFGPEVCITHQGE----- 323
 DB 1177 EBEAKTHEAQIQEMRCKHSAVEELAEQLEQTQKRVKANLEKAKQT-LENERGELANEVKV 1235

QY 324 -----KISANENSLAVRSTPAEDSDSDSQVSEVQVQVHPKPLSPDSRAS-- 368
 DB 1236 LLOQKGSEHKKRKYEAQLOELQVKNFE-----RVRTLADKVTKLQVELONVTOLL 1289
 QY 369 SLSESSPPKAMKFOAPARETCVCOKTVYPMERLLANQQVHFHSCPRCSYCNKLSIGT 428
 DB 1290 SOSDSKSKLTDFSA-----LESQLOQTQELQEE-----ROKLSUST 1329
 QY 429 YASLHGRIYCKPHFNQLFKSKGNDE--GFGHRPHKDLWASKNENEERILERPQALANARE 486
 DB 1330 -----KLVQVEDEKNSFREOLEEBEAEKHNLEKQIATLHAQVADMKK 1371
 QY 487 TPHSPGVEDAPIAKVGVLAASMEAKASSQEKEDKPAETKKLRIAMPPPTTELGGSSGSALE 546
 DB 1372 K-----MEDS-----VGCLETAEEVKRKLQKLEGLSQRHEBKVAAYD----- 1409
 QY 547 EGIKMSKPKWPPDEDEISKPEVPDVLKLRSS--SLKERSRPF-----TVAASF 597
 DB 1410 ---KLEKTKTSLQOEL-----DQLLVLDHQROSACNLEKKQKKFQOLLAEEKTSISKY 1460
 QY 598 QSTSVKSPKTVSPPIRKWSMSEQSESVGGRVAE-----RKOVENAKSKKNGNVG 649
 DB 1461 ABEEDRAEAEAREKETKALSARALEEAMBOK-AELERLNKQFTEWEDLMSSK--DDVG 1517
 QY 650 KTTW----- 653
 DB 1518 KSVHELEKSKREALSQVVEEMKTLQEELEDELEQATEDAKLRLEVNLMQAKQAFERDQGRD 1577
 QY 554 -ONKESGKETCKRSKEGHS-----LEMENENLVENGADSDDEDSNFL 694
 DB 1578 EQSEKKKQLVRQVREMEAELEDERKORSMAVAARKKLEMDLKDLKLEAHIDSANKNRDEAI 1637
 QY 695 KQOSPQEPKSLNWSFVQNTFA--EFTTO-NQKSDQVELWEGEVV---KELSVEEQIKR 748
 DB 1638 KQLRKLAQMKDCMRLELDDTRASREEILAQAKENEKKLKSMEAEMLQLEELAAAEAKR 1697
 QY 749 NRYVDEDE 756
 DB 1698 QAQQRDE 1705

Search completed: January 6, 2004, 09:48:25

Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:43:29 ; Search time 48 Seconds

(without alignments)
2509.864 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSFPNRQWTSLSLRVTA.....LSVEQIKRNYDEDEDEE 759

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq 19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3927	100.0	759	21	AA197286 Lipid associated p
2	3927	100.0	759	22	ABUS2869 Human transcriptio
3	3927	100.0	759	22	AAM39013 Human polypeptide
4	3925	99.9	759	22	AAB67701 Human EPLIN (epith
5	3909.5	99.6	760	21	AAB42934 Human ORFX-ORF2698
6	3083.5	78.5	600	22	AAB67700 Human EPLIN (epith
7	2458	62.6	471	20	AA148487 Human breast tumou
8	2394	61.0	457	22	AAB95477 Human protein sequ
9	2378	60.6	457	23	AAU10979 Human Sterol regul

10	2011	51.2	476	22	ABG08559 Novel human diagno
11	1548	39.4	299	21	AAB34206 Human secreted pro
12	1473	37.5	301	23	ABB56420 Human cancer suppl
13	1254	31.9	243	21	AAB54159 Human pancreatic c
14	1158	29.5	232	21	AAB51883 Human secreted pro
15	1068	27.2	299	21	AAB34205 Gene 41 human secr
16	889.5	22.7	235	20	AAY73916 Human prostate tum
17	815	20.8	337	22	ABG24050 Novel human diagno
18	781	19.9	234	21	AAB51882 Gene 5 human secre
19	583	14.8	152	22	AAM40799 Human polypeptide
20	521.5	13.3	547	22	AU20575 Human secreted pro
21	521.5	13.3	547	22	AU20575 Human secreted pro
22	521.5	13.3	547	22	AU20575 Human secreted pro
23	464.5	11.8	527	23	ABP43758 Human breast tumou
24	464.5	11.8	527	23	ABP43758 Human breast tumou
25	393	10.0	88	21	AAG02073 Human protein #5 c
26	385.5	9.8	171	22	ABG04347 Novel human diagno
27	381	9.7	135	22	ABG12836 Novel human diagno
28	363	9.2	158	21	AAY79176 Haematopoietic ste
29	357	9.1	127	21	AAY94447 Human inflammation
30	357	9.1	173	23	ABG06053 Human NS protein s
31	357	9.1	173	23	ABG06054 Human NS protein s
32	357	9.1	173	23	ABG06055 Human NS protein s
33	357	9.1	173	23	ABG06056 Human NS protein s
34	357	9.1	173	23	ABG06057 Human NS protein s
35	357	9.1	173	23	ABG06058 Human NS protein s
36	354	9.0	111	22	ABG12835 Novel human diagno
37	325	8.3	217	22	AU20650 Human secreted pro
38	247	6.3	331	22	ABG08554 Novel human diagno
39	232.5	5.9	1793	22	ABG59613 Drosophila melanog
40	231	5.9	134	21	AAB32979 Pinus radiata tran
41	231	5.9	3111	22	ABG60327 Drosophila melanog
42	229	5.8	193	21	AAB32984 Pinus radiata tran
43	226	5.8	993	22	AAM39437 Human polypeptide
44	226	5.8	1312	22	ABG50244 Human transcriptio
45	223	5.7	1312	21	AAY76884 Retinoblastoma bin

ALIGNMENTS

RESULT 1

AA197286

ID AA197286 standard; Protein; 759 AA.

AC AA197286;

XX 03-JAN-2001 (first entry)

XX Lipid associated protein (LIPAP) 2766980CD1.

KW Lipid associated protein; LIPAP; treatment; prophylaxis; agonist;
KW antagonist; antibody; cardiovascular disease; neurological disease;
KW gastrointestinal disease; lipid metabolism; detection;
KW amplification; monitoring; hybridisation; antisense; triplex;
KW ribozyme; screening; immunoassay.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 15
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 19
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 29
FT Modified-site /note= "Potential glycosylation site"
FT Modified-site 48
FT Modified-site /note= "Potential glycosylation site"
FT Modified-site 55
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 59
FT Modified-site /note= "Potential phosphorylation site"
FT Modified-site 68
FT Modified-site /note= "Potential glycosylation site"

FT	Modified-site	658	/note= "Potential phosphorylation site"
FT	Modified-site	662	/note= "Potential phosphorylation site"
FT	Modified-site	686	/note= "Potential phosphorylation site"
FT	Modified-site	698	/note= "Potential phosphorylation site"
FT	Modified-site	706	/note= "Potential glycosylation site"
FT	Modified-site	709	/note= "Potential phosphorylation site"
FT	Modified-site	714	/note= "Potential phosphorylation site"
FT	Modified-site	741	/note= "Potential phosphorylation site"
XX			
PN	W0200049043-A2.		
XX			
PD	24-AUG-2000.		
XX			
PF	18-FEB-2000; 2000WO-US04160.		
XX			
PR	19-FEB-1999; 99US-0120703.		
PR	08-JUL-1999; 99US-0142762.		
XX			
PA	(INCY-) INCYTE PHARM INC.		
XX			
PI	Tang YT, Hillman JL, Yue H, Azimzai Y, Baughn MR, Tran B;		
XX			
DR	WPI; 2000-549264/50.		
DR	N-PSDB; AAA53826.		
XX			
PT	New human lipid-associated proteins, nucleic acids, and antibodies,		
PT	useful for diagnosis, treatment and prevention of e.g. cardiovascular		
PT	disease		
XX			
PS	Claim 1; Page 76-77; 93pp; English.		
XX			
CC	Lipid-associated proteins (LIPAP) can be used for treating or		
CC	preventing disorders associated with decreased expression of LIPAP,		
CC	for screening for agonists or antagonists of LIPAP, and to raise		
CC	specific antibodies. Antagonists and antagonists of LIPAP are useful		
CC	for treating diseases associated with reduced or increased levels of		
CC	LIPAP, e.g. cardiovascular, neurological and gastrointestinal		
CC	diseases and disorders of lipid metabolism. Fragments of the nucleic		
CC	acid encoding LIPAP are useful for detection of full length coding		
CC	sequences, in hybridization and/or amplification assays or for		
CC	diagnosis or monitoring. Nucleotides encoding LIPAP are used		
CC	to screen for compounds that specifically modify LIPAP expression,		
CC	for recombinant production of LIPAP, in gene therapy, as a source of		
CC	therapeutic antisense, triplex-forming, or ribozyme agents and for		
CC	genomic mapping. Antibodies to the proteins are used for diagnosis		
CC	and monitoring of LIPAP-associated disease by immunoassay, as		
CC	antagonists, in competitive drug screens and for affinity		
CC	purification of natural LIPAP.		
XX			
SQ	Sequence 759 AA.		
SQ			

	Query Match	100.0.0;	Score 3927;	DB 21;	Length 759;
	Best Local Similarity	100.0.0;	Pred. No. 6.1e-299;		
	Matches 759;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MESSPNNRRQWTSLSURVTAKLSLSLVKNKNSAIVEIFSKYQKAAEETNMKKRSNTNL	60		
Db	1	MESSPNNRRQWTSLSURVTAKLSLSLVKNKNSAIVEIFSKYQKAAEETNMKKRSNTNL	60		
QY	61	SOHFRKGTTLVLKKKWENPGLGAESHTDSLRNSSTEIRHRADHPAEVTSHAASGAKAQD	120		
Db	61	SOHFRKGTTLVLKKKWENPGLGAESHTDSLRNSSTEIRHRADHPAEVTSHAASGAKAQD	120		
QY	121	EGCITHSPRSRLRPPPEALVQGRYPHTKDGEDLKHDSHSTESKMCNCLGESRHEVKEISEIN	180		

Db 121 BEQIHPRRLRSPALVQGRYPHIKDGEDLDKDHSTESKMKMNCIGESRHEVEKSEISEN 180
Qy 181 TDASGKIEKYNVPLNRLKMFKEGPTQTKILRAQSRASGRKISENSYSLDDLEIGPGQ 240
Db 181 TDASGKIEKYNVPLNRLKMFKEGPTQTKILRAQSRASGRKISENSYSLDDLEIGPGQ 240
Qy 241 LSSSTFDSEKNESSRNLELPRLSSETSISKDRMAKYAAVSKOSSSTNYTNELKASGEIKI 300
Db 241 LSSSTFDSEKNESSRNLELPRLSSETSISKDRMAKYAAVSKOSSSTNYTNELKASGEIKI 300
Qy 301 HMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVKEVQPVHPKP 360
Db 301 HMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVKEVQPVHPKP 360
Qy 361 LSPDSRASSLSSESSPPKAMKXFOAPARETCVEQKTVYPMERLLANQOVFHSFCRCSYC 420
Db 361 LSPDSRASSLSSESSPPKAMKXFOAPARETCVEQKTVYPMERLLANQOVFHSFCRCSYC 420
Qy 421 NNKLSLGTYSALHGRIYCKPHFNQLFKSGKGYDEGFGHRPHKDLWASKNENEELERPAQ 480
Db 421 NNKLSLGTYSALHGRIYCKPHFNQLFKSGKGYDEGFGHRPHKDLWASKNENEELERPAQ 480
Qy 481 LANARETPHSPGVEDAPIAKVGVLAASMEAKASQOEKEDKPAETKKLRIANPPPTLGS 540
Db 481 LANARETPHSPGVEDAPIAKVGVLAASMEAKASQOEKEDKPAETKKLRIANPPPTLGS 540
Qy 541 SGSALEEGIKMSKPKWPPEDISKEPEVEDVDLDLKKLRRSSSLKERSRPTTVAASTQST 600
Db 541 SGSALEEGIKMSKPKWPPEDISKEPEVEDVDLDLKKLRRSSSLKERSRPTTVAASTQST 600
Qy 601 SVKSPKTVSPPIRKGWSMESEESVGGVRAERKOVENAKASKKNGVNGKTTWQNKESKG 660
Db 601 SVKSPKTVSPPIRKGWSMESEESVGGVRAERKOVENAKASKKNGVNGKTTWQNKESKG 660
Qy 661 ETGKRSKEGHSLEMEENLVENGADSDDEDDNSFLKQSPQEPKSLNWSSFVDNTFAEFT 720
Db 661 ETGKRSKEGHSLEMEENLVENGADSDDEDDNSFLKQSPQEPKSLNWSSFVDNTFAEFT 720
Qy 721 TQNKSQDVELWEGEVVKELSVEEQIKNRYVDEDEDEE 759
Db 721 TQNKSQDVELWEGEVVKELSVEEQIKNRYVDEDEDEE 759

RESULT 2
ABUS2869
ID ABUS2869 standard; Protein; 759 AA.
AC ABUS2869;
XX
DT 14-APR-2003 (first entry)
XX
DE Human transcription factor from DKFzphutcl_18c19.
XX
KW Human; gene therapy; vaccine; disease treatment; detection.
XX
OS Homo sapiens.
XX
PN WO200112659-A2.
XX
XX 22-FEB-2001.
XX
PF 18-AUG-2000; 2000WO-IB01496.
XX
PR 18-AUG-1999; 99US-0149499.
XX
PR 28-SEP-1999; 99US-0156503.
XX
PA (GSHU-) GERMAN HUMAN GENOME PROJECT.
XX
XX Wiemann S;
XX
XX WPI; 2001-327840/34.
XX
XX N-PSDB; ABX71307.
XX

PT Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies -
XX
PS Claim 21; Page 450-451; 1095pp; English.
XX
CC This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a
CC polypeptide described in the disclosure of the invention.
XX

Sequence 759 AA;

Query Match 100.0%; Score 3927; DB 22; Length 759;
Best Local Similarity 100.0%; Pred. No. 6.1e-299;
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MESSPNNRRRMTSLSRVTAKELSLVKNKSSAIVEIFSKYKAAABETNMEKRSNTENL 60
Db 1 MESSPNNRRRMTSLSRVTAKELSLVKNKSSAIVEIFSKYKAAABETNMEKRSNTENL 60
Qy 61 SQHFRKCTLTVLKKWENPCLGAESHTDSLRSNSTEIRHRADHPPAEVTSHAASGAKAQ 120
Db 61 SQHFRKCTLTVLKKWENPCLGAESHTDSLRSNSTEIRHRADHPPAEVTSHAASGAKAQ 120
Qy 121 EQIHPRSLRSPPEALVQGRYPHIKDGEDLDKDHSTESKMKMNCIGESRHEVEKSEISEN 180
Db 121 EQIHPRSLRSPPEALVQGRYPHIKDGEDLDKDHSTESKMKMNCIGESRHEVEKSEISEN 180
Qy 181 TDASGKIEKYNVPLNRLKMFKEGPTQTKILRAQSRASGRKISENSYSLDDLEIGPGQ 240
Db 181 TDASGKIEKYNVPLNRLKMFKEGPTQTKILRAQSRASGRKISENSYSLDDLEIGPGQ 240
Qy 241 LSSSTFDSEKNESSRNLELPRLSSETSISKDRMAKYAAVSKOSSSTNYTNELKASGEIKI 300
Db 241 LSSSTFDSEKNESSRNLELPRLSSETSISKDRMAKYAAVSKOSSSTNYTNELKASGEIKI 300
Qy 301 HMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVKEVQPVHPKP 360
Db 301 HMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVKEVQPVHPKP 360
Qy 361 LSPDSRASSLSSESSPPKAMKXFOAPARETCVEQKTVYPMERLLANQOVFHSFCRCSYC 420
Db 361 LSPDSRASSLSSESSPPKAMKXFOAPARETCVEQKTVYPMERLLANQOVFHSFCRCSYC 420
Qy 421 NNKLSLGTYSALHGRIYCKPHFNQLFKSGKGYDEGFGHRPHKDLWASKNENEELERPAQ 480
Db 421 NNKLSLGTYSALHGRIYCKPHFNQLFKSGKGYDEGFGHRPHKDLWASKNENEELERPAQ 480
Qy 481 LANARETPHSPGVEDAPIAKVGVLAASMEAKASQOEKEDKPAETKKLRIANPPPTLGS 540
Db 481 LANARETPHSPGVEDAPIAKVGVLAASMEAKASQOEKEDKPAETKKLRIANPPPTLGS 540
Qy 541 SGSALEEGIKMSKPKWPPEDISKEPEVEDVDLDLKKLRRSSSLKERSRPTTVAASFQST 600
Db 541 SGSALEEGIKMSKPKWPPEDISKEPEVEDVDLDLKKLRRSSSLKERSRPTTVAASFQST 600
Qy 601 SVKSPKTVSPPIRKGWSMESEESVGGVRAERKOVENAKASKKNGVNGKTTWQNKESKG 660
Db 601 SVKSPKTVSPPIRKGWSMESEESVGGVRAERKOVENAKASKKNGVNGKTTWQNKESKG 660
Qy 661 ETGKRSKEGHSLEMEENLVENGADSDDEDDNSFLKQSPQEPKSLNWSSFVDNTFAEFT 720
Db 661 ETGKRSKEGHSLEMEENLVENGADSDDEDDNSFLKQSPQEPKSLNWSSFVDNTFAEFT 720
Qy 721 TQNKSQDVELWEGEVVKELSVEEQIKNRYVDEDEDEE 759
Db 721 TQNKSQDVELWEGEVVKELSVEEQIKNRYVDEDEDEE 759

RESULT 3	QY	1	MESSPPNRRQWTSLSLRTAKELSLVNNKSSAIVEIFSKYQKAAEETNNKKRSNTENL	60
AAM39013	DB	1	MESSPPNRRQWTSLSLRTAKELSLVNNKSSAIVEIFSKYQKAAEETNNKKRSNTENL	60
ID AAM39013 standard; Protein; 759 AA.	QY	61	SOHERKGTLTVLKKWENPGLGAESHTDSLRSNSTEIRHRADHPPAEVTSHAASGAKAQD	120
AC AAM39013;	DB	61	SOHERKGTLTVLKKWENPGLGAESHTDSLRSNSTEIRHRADHPPAEVTSHAASGAKAQD	120
XX	QY	121	EQIHRPSRLRSPPPEALVQGRYPHIKDGEDLKHSTESKMWENCLGESRHEVEKSEISEN	180
DT 22-OCT-2001 (first entry)	DB	121	EQIHRPSRLRSPPPEALVQGRYPHIKDGEDLKHSTESKMWENCLGESRHEVEKSEISEN	180
DE Human polypeptide SEQ ID NO 2159.	QY	181	TDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRSASGRKISSENSYLLDLEIGPGQ	240
XX	DB	181	TDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRSASGRKISSENSYLLDLEIGPGQ	240
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;	QY	241	LSSSTFDSKESRRNLELPRLETSIKDRMAKYQAAVSKQSSSTNTYNELKASGGBIKI	300
KW peripheral nervous system; neuropathy; central nervous system; CNS;	DB	241	LSSSTFDSKESRRNLELPRLETSIKDRMAKYQAAVSKQSSSTNTYNELKASGGBIKI	300
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	QY	301	HKMEQKENVPPGPEVCITHOGEKISANENSLAVRSTPAEDDSDRSOVKSEVQPVHPKP	360
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	DB	301	HKMEQKENVPPGPEVCITHOGEKISANENSLAVRSTPAEDDSDRSOVKSEVQPVHPKP	360
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	QY	361	LSPDSRASSLSSESPPKAMKKFOAPARETCVECKTYPMERLLANQOVFHSICFRCSYC	420
KW leukaemia.	DB	361	LSPDSRASSLSSESPPKAMKKFOAPARETCVECKTYPMERLLANQOVFHSICFRCSYC	420
XX Homo sapiens.	QY	421	NNKLSLGTYSALHGRITYCKPHFNQLFKSKNGYDEGFGHRPHKOLWASKNENEELERPAQ	480
OS	DB	421	NNKLSLGTYSALHGRITYCKPHFNQLFKSKNGYDEGFGHRPHKOLWASKNENEELERPAQ	480
XX WO200153312-A1.	QY	481	LANARETFPHSPGVEDAPIAKGVGLAASMEAKASQOQEKEDKPAETKKLR.AWPPPTLGS	540
PN	DB	481	LANARETFPHSPGVEDAPIAKGVGLAASMEAKASQOQEKEDKPAETKKLR.AWPPPTLGS	540
PD 26-JUL-2001.	QY	541	SGSALLEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPTTVAASFQST	600
XX	DB	541	SGSALLEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPTTVAASFQST	600
PF 26-DEC-2000; 2000WO-US34263.	QY	601	SVKSPKTVSPPIRKGMSESQSESVGGRVAERKQVENAKSKNGNVGKTTWQNKESKG	660
PR	DB	601	SVKSPKTVSPPIRKGMSESQSESVGGRVAERKQVENAKSKNGNVGKTTWQNKESKG	660
XX 21-JAN-2000; 2000US-0488725.	QY	661	ETGKRSKEGHSLEMNENLVENGADSDDEDDNSFLKQSPQEPKSLNWSFVONTFAFEFT	720
PR	DB	661	ETGKRSKEGHSLEMNENLVENGADSDDEDDNSFLKQSPQEPKSLNWSFVONTFAFEFT	720
PR 25-APR-2000; 2000US-0552317.	QY	721	TQNXQSDQVLEWGEVVKELSVVEQIKENRYDYDEDEE 759	
PR	DB	721	TQNXQSDQVLEWGEVVKELSVVEQIKENRYDYDEDEE 759	
PR 09-JUL-2000; 2000US-0598042.	QY		RESULT 4	
PR	DB		AAB67701	
PR 19-JUL-2000; 2000US-0620312.	QY	ID	AAB67701 standard; Protein; 759 AA.	
PR	DB	XX	AAB67701;	
PR 03-AUG-2000; 2000US-0653450.	QY	AC	AAB67701;	
PR	DB	XX	11-JUN-2001 (first entry)	
PR 14-SEP-2000; 2000US-0662191.	QY	DT	Human EPLIN (epithelial protein lost in neoplasm)-beta isoform.	
PR	DB	XX	Human EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;	
PR 19-OCT-2000; 2000US-0693036.	QY	KW	EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;	
PR	DB	XX	gene therapy; cancer.	
PR 29-NOV-2000; 2000US-0727344.	QY	OS	Homo sapiens.	
XX	DB	XX	Key	Location/Qualifiers
(HYSE-) HYSEQ INC.	QY	FT	Misc-difference 278	
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	DB	FT	Misc-difference 344	
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;				
Zhao QA, Zhou P, Goodrich R, Drmanac RT;				
WPI; 2001-442253/47.				
N-PSDB; AAI58169.				
Novel nucleic acids and polypeptides, useful for treating disorders				
such as central nervous system injuries -				
Example 4; SEQ ID NO 2159; 10078pp; English.				
The invention relates to human nucleic acids (AAI57798-AAI61369) and				
the encoded polypeptides (AAM39642-AA42213) with nootropic,				
immunosuppressant and cytostatic activity. The polynucleotides are useful				
in gene therapy. A composition containing a polypeptide or polynucleotide				
of the invention may be used to treat diseases of the peripheral nervous				
system, such as peripheral nervous injuries, peripheral neuropathy and				
centralised neuropathies and central nervous system diseases, such as				
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic				
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the				
utilisation of the activities such as: Immune system suppression,				
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic				
and thrombolytic activity, cancer diagnosis and therapy, drug screening,				
C.N.S disorders.				
Note: The sequence data for this patent did not form part of the printed				
specification.				
Sequence 759 AA;				
Query Match				
Best Local Similarity 100.0%; Score 3927; DB 22; Length 759;				
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				

FT Misc-difference 495 /note= "Arg encoded by CCAGGT"
FT FT /note= "Asp encoded by AAT"
FT Misc-difference 511 /note= "Lys encoded by AGG"
FT Misc-difference 622 /note= "Ser encoded by AAT"
FT Misc-difference 658 /note= "Ser encoded by TTT"
FT Misc-difference 679 /note= "Ile encoded by TTT"
FT Misc-difference 698 /note= "Ser encoded by TTT"
FT Misc-difference 704 /note= "Ser encoded by TTT"
XX
PN WO200118019-A1.
XX
PD 15-MAR-2001.
XX
XX
PF 08-SEP-2000; 2000WO-US24689.
XX
PR 08-SEP-1999; 99US-0153024.
XX
PA (REGC) UNIV CALIFORNIA.
XX
XX Chang DD, Maul RS;
PI
XX WPI; 2001-244555/25.
XX N-PSDB; AAF55697.
DR
DR New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
PT prognostic and therapeutic applications over the course of cell
PT proliferative disorders associated with EPLIN
XX
XX Claim 1; Page 44; 59pp; English.
XX
CC The present sequence represents a human EPLIN (epithelial protein lost in
CC neoplasm)-beta isoform. The specification also describes EPLIN-alpha.
CC EPLIN is a tumour suppressor protein, whose expression is altered in
CC multiple common human tumour types. EPLIN nucleic acids and proteins are
CC used in screening assays to detect molecules that specifically bind to
CC EPLIN nucleic acids, proteins or derivatives and thus have potential use
CC as agonist or antagonist of EPLIN, in particular molecules that affect
CC cell proliferation. Thus the assays are useful for screening molecules
CC with potential utility as anticancer drugs or lead compounds for drug
CC development. EPLIN nucleic acids, proteins are useful for detecting a
CC cell proliferative disorder in a subject. EPLIN polynucleotides are
CC useful in gene therapy techniques. EPLIN is useful as a marker that
CC can be diagnostically, prognostically and therapeutically used over
CC the course of a cell proliferative disorder associated with EPLIN.
XX
SQ Sequence 759 AA;
Query Match 99.9%; Score 3925; DB 22; Length 759;
Best Local Similarity 99.9%; Pred. No. 8.8e-299;
Matches 758; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Y 1 MESSPPNRRQWTSLSLRVAKELSLVKNKSSAIVEIFSKYKAAETNNKRSNTENL 60
Db 1 MESSPPNRRQWTSLSLRVAKELSLVKNKSSAIVEIFSKYKAAETNNKRSNTENL 60
Y 61 SQHFRKGTTLVKKWENPGILGAESHTDSLRSNSTEIRHRADHPPEVTSAAASGAKADQ 120
Db 61 SQHFRKGTTLVKKWENPGILGAESHTDSLRSNSTEIRHRADHPPEVTSAAASGAKADQ 120
Y 121 EQHPRSRRLSPPEALVQGRYPHIKQDGLKHSTESKKWENCLGESRHEVEKSEISEN 180
Db 121 EQHPRSRRLSPPEALVQGRYPHIKQDGLKHSTESKKWENCLGESRHEVEKSEISEN 180
Y 181 TDASGKTEKYNVPLNRLKWMFEKCEPTQTTLKRAQSASGRKISENSYSLDLEIGPGQ 240
Db 181 TDASGKTEKYNVPLNRLKWMFEKCEPTQTTLKRAQSASGRKISENSYSLDLEIGPGQ 240

QY 241 LSSSTFDSSEKXNRRNLELPRLSETSIKORMAKYQAAVSKQSSSTNYTNELKASGGEIKI 300
Db 241 LSSSTFDSSEKXNRRNLELPRLSETSIKORMAKYQAAVSKQSSSTNYTNELKASGGEIKI 300
QY 301 HKMEQKENVPPGPEVCITHOEGEKISANENSLAVRSTPAEDDSRDQVSEVQPVHPKP 360
Db 301 HKMEQKENVPPGPEVCITHOEGEKISANENSLAVRSTPAEDDSRDQVSEVQPVHPKP 360
QY 361 LSPDSRASSLSSESPPKAMKKFQAPARETCVEQKTVYPMERLLANQOVFHI5CFRC5YC 420
Db 361 LSPDSRASSLSSESPPKAMKKFQAPARETCVEQKTVYPMERLLANQOVFHI5CFRC5YC 420
QY 421 NNKLSLGTASLHGRYICKPHFNQLFKSKGNVDEGFGHPRPHKDLWASKNENEBILERRPAQ 480
Db 421 NNKLSLGTASLHGRYICKPHFNQLFKSKGNVDEGFGHPRPHKDLWASKNENEBILERRPAQ 480
QY 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRITAWPPPTLGS 540
Db 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRITAWPPPTLGS 540
QY 541 SGSALBEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKKLRRSSSLKERSRPFVVAASFQST 600
Db 541 SGSALBEGIKMSKPKWPPPEDEISKPEVPEDVDLDLKKLRRSSSLKERSRPFVVAASFQST 600
QY 601 SVKSPKTVSPPIRKGWSMGESESVGGRVAERKQVENAKASKKNGNVGKTTWQNKESKG 660
Db 601 SVKSPKTVSPPIRKGWSMGESESVGGRVAERKQVENAKASKKNGNVGKTTWQNKESKG 660
QY 661 ETCKRSKEGSHLEMEENENENENGADSDDEDNSFLKQSPQEPKSLNWSFVDNTFAEFT 720
Db 661 ETCKRSKEGSHLEMEENENENENGADSDDEDNSFLKQSPQEPKSLNWSFVDNTFAEFT 720
QY 721 TQNKQSQDVELWEGEVVVEEQIKRNRYYDEDEDEE 759
Db 721 TQNKQSQDVELWEGEVVVEEQIKRNRYYDEDEDEE 759
RESULT 5
AAB42934
ID AAB42934 standard; Protein; 760 AA.
XX
AC AAB42934;
XX
DT 08-FEB-2001 (first entry)
DE Human ORFX ORF2698 polypeptide sequence SEQ ID NO:5396.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
XX N-PSDB; AAC77143.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 4580-4582; 5507pp; English.
XX
XX AAC74446 to AAC7606 encode the proteins given in AAC40237 to AAC43397,
CC which represent the human ORF open reading frames 1 to 3161. The ORF
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; neurotropic; immunoprotective;
CC osteopathic; anticonvulsant; antiarthritic; neurosuppressive;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORF-associated disorder. The
CC nucleic acids can be used to express ORF proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 760 AA;

Query Match 99.6%; Score 3909.5; DB 21; Length 760;
Best Local Similarity 99.7%; Pred. No. 1.4e-297;
Matches 756; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MESSPNRRQWTSLSRVTAKELSLVNKKSSAIVEIFSKYQAAETNNEKRSNTNL 60
DB 1 MESSPNRRQWTSLSRVTAKELSLVNKKSSAIVEIFSKYQAAETNNEKRSNTNL 60

QY 61 SQHFRKGTTLVLKKWENPGLGAESHTSLRNSSTBIHRADHPPEAVTSHAASAKAQD 120
DB 61 SQHFRKGTTLVLKKWENPGLGAESHTSLRNSSTBIHRADHPPEAVTSHAASAKAQD 120

QY 121 EQQIHPRSLRSPPEALVQGRYPHIKDGEDLKHSTESKMNCLGESRHEVEKSEISEN 180
DB 121 EQQIHPRSLRSPPEALVQGRYPHIKDGEDLKHSTESKMNCLGESRHEVEKSEISEN 180

QY 181 TDASGKIEKYNVPLNRLKMMFKGEPTQTKILRAQSRASGRKISSENSYSLDDELEIGPQQ 240
DB 181 TDASGKIEKYNVPLNRLKMMFKGEPTQTKILRAQSRASGRKISSENSYSLDDELEIGPQQ 240

QY 241 LSSSTFDSEKNSRRNLELPRLSSETSIKORMAKYQAAVSKQSSSTNTYNELKASGEIKI 300
DB 241 LSSSTFDSEKNSRRNLELPRLSSETSIKORMAKYQAAVSKQSSSTNTYNELKASGEIKI 300

QY 301 HKMEQENVPVPGPEVITIQEGEKISANENSIAVRSTPAEDDS-RDSQYKSEVQPVHPK 359
DB 301 HKMEQENVPVPGPEVITIQEGEKISANENSIAVRSTPAEDDSPGDSQYKSEVQPVHPK 360

QY 360 PLSPDSRASLSSESPKAMKFOAPARETCVCKQTVYPMERLLANQOVFIHSCFRCSY 419
DB 361 PLSPDSRASLSSESPKAMKFOAPARETCVCKQTVYPMERLLANQOVFIHSCFRCSY 420

QY 420 CNNKLSLGTYSASLHGRYCKPHFNQLFKSGNYDEGFGHRPHKDLWASKNENEILERP 479

DB 421 CNNKLSLGTYSASLHGRYCKPHFNQLFKSGNYDEGFGHRPHKDLWASKNENEILERP 480
QY 480 QLANARETPHSPGVEDAPIAKVGVLAASMEAKASQOEKEDKPAETKKLRIAWPPPTTEL 539
DB 481 QLANARETPHSPGVEDAPIAKVGVLAASMEAKASQOEKEDKPAETKKLRIAWPPPTTEL 540
QY 540 SSGSALESGIKMSKPKWPPPEDEISKPEVPEDVDLDLKKLRSSSLKRSRPFVVAASFQS 599
DB 541 SSGSALESGIKMSKPKWPPPEDEISKPEVPEDVDLDLKKLRSSSLKRSRPFVVAASFQS 600
QY 600 TSVKSPKTVSPPIRKGMMSQSEESVGGRAERKQVENAKASKKNGVNGKTTWQNKESK 659
DB 601 TSVKSPKTVSPPIRKGMMSQSEESVGGRAERKQVENAKASKKNGVNGKTTWQNKESK 660
QY 660 GETGKRSKEGHSLEMNENLVENGADSDDEDDNSFLKQSPQEPKSLNWSVVDNTFAEEF 719
DB 661 GETGKRSKEGHSLEMNENLVENGADSDDEDDNSFLKQSPQEPKSLNWSVVDNTFAEEF 720
QY 720 TTQNKQSDQVLEWGEVVKLSVEEQIKRNYRYYDEDEE 759
DB 721 TTQNKQSDQVLEWGEVVKLSVEEQIKRNYRYYDEDEE 760

RESULT 6
AAB67700
ID AAB67700 standard; Protein; 600 AA.
XX
AC AAB67700;
XX
DT 11-JUN-2001 (first entry)
XX Human EPLIN (epithelial protein lost in neoplasm)-alpha isoform.
DE Human; EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;
KW EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;
KW gene therapy; cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 118
FT Misc-difference /note= "Val encoded by GCG"
FT Misc-difference 336
FT Misc-difference /note= "App encoded by AAT"
FT Misc-difference 352
FT Misc-difference /note= "Lys encoded by AGG"
FT Misc-difference 463
FT Misc-difference /note= "Ser encoded by AAT"
FT Misc-difference 499
FT Misc-difference /note= "Ser encoded by TTT"
FT Misc-difference 520
FT Misc-difference /note= "Ile encoded by TTT"
FT Misc-difference 539
FT Misc-difference /note= "Ser encoded by TTT"
FT Misc-difference 545
FT Misc-difference /note= "Ser encoded by TTT"
XX
PN WO200118019-A1.
XX
XX 15-MAR-2001.
XX
XX 08-SEP-2000; 2000WO-US44689.
XX
XX 08-SEP-1999; 99US-0153024.
XX (REGC) UNIV CALIFORNIA.
XX Chang DD, Maul RS;
XX WPI; 2001-244555/25.
XX N-PSDB; AAF55696.
XX

PT New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
PT prognostic and therapeutic applications over the course of cell
PT proliferative disorders associated with EPLIN -
XX
PS Claim 1; Page 43; 59pp; English.

The present sequence represents a human EPLIN (epithelial protein lost in neoplasm)-alpha isoform. The specification also describes EPLIN-beta. EPLIN is a tumour suppressor protein, whose expression is altered in multiple common human tumour types. EPLIN nucleic acids and proteins are used in screening assays to detect molecules that specifically bind to EPLIN nucleic acids, proteins or derivatives and thus have potential use as agonist or antagonist of EPLIN, in particular molecules that affect cell proliferation. Thus the assays are useful for screening molecules with potential utility as anticancer drugs or lead compounds for drug development. EPLIN nucleic acids, proteins are useful for detecting a cell proliferative disorder in a subject. EPLIN polynucleotides are useful in gene therapy techniques. EPLIN is useful as a marker that can be diagnostically, prognostically and therapeutically used over the course of a cell proliferative disorder associated with EPLIN.

Sequence 600 AA;

Query Match	78.5%	Score 3083.5;	DB 22;	Length 600;
Best Local Similarity	99.5%	Pred. No. 6.4e-233;		
Matches 597; Conservative	1;	Mismatches 1;	Indels 1;	

QY	161	MENCLGSRHVEYKSEISENTDASGKIEKYNVPLNRLKOMFEKGEPTQTKILRAQSRAS	220
		=====	
Db	1	MENCLGSRHVEYKSEISENTDASGKIEKYNVPLNRLKOMFEKGEPTQTKILRAQSRAS	60

Oy		221 GRKISENSYSLDLEIGPQGLSSTFDSEKNESRRNLEIPRLSTSICKORMAKYCQAASVK 280
Dd		61 GRKISENSYSLDLEIGPQGLSSTFDSEKNESRRNLEIPRLSTSICKORMAKYCQAASVK 120

Qy	QSSSTVNTNELKASGGGEIKIHKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAE	340
281	QSSSTVNTNELKASGGGEIKIHKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAE	340
121	QSSSTVNTNELKASGGGEIKIHKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAE	180
Db	QSSSTVNTNELKASGGGEIKIHKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAE	180

QY 341 DDS -RDSQVKSEVQQVHPKPLSPDSSRASSLESSPPKAMKKFOAPARETCVECKTYP 399
||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Dd 181 DDPGDSQVKSEVQQVHPKPLSPDSSRASSLESSPPKAMKKFOAPARETCVECKTYP 240

Qy 400 MERLLANQOVPHISCFRCSYCNKKLSLGTYSALHGRIYCKPHFNQLFKSGNGYDEGFGHR 459
|||||
241 MERLLANQOVPHISCFRCSYCNKKLSLGTYSALHGRIYCKPHFNQLFKSGNGYDEGFGHR 300
Db

[illegible]

QY 520 DKPAETKKLRIAWPPPTLGGSSALEEGIKMSKPWPPEDEISKPEVEDVDJLKKLR 579
|||
361 DKPAETKKLRIAWPPPTLGGSSALEEGIKMSKPWPPEDEISKPEVEDVDJLKKLR 420
Db

QY 580 RSSSLAERSRPPTVAASFQSTSVKSPKTVSPPIRKGWNSSEQSEESVGRVAERKQVENA 639
 |||||
 Db 421 RSSSLAERSRPPTVAASFQSTSVKSPKTVSPPIRKGWNSSEQSEESVGRVAERKQVENA 480

QY 640 KASKKNGVGTWQNKESKGETGKRSKEGHSLEMEMENINVENGADSDDEDDNSFLKQOSP 699
DB 481 KASKKNGVGTWQNKESKGETGKRSKEGHSLEMEMENINVENGADSDDEDDNSFLKQOSP 540

QY 700 QEPKSLNWSSFFVDNTFAAEFTTQKSDQVLMWEGEVVKELSVEEQIKRNRYYDEDEDE 759

Db 541 QEPKSLNWSSFFVDNTFAAEFTTQKSDQVLMWEGEVVKELSVEEQIKRNRYYDEDEDE 600

RESULT 7
AA48487
ID AA48487 standard; Protein: 471 AA.

AA AC AAY48487;

XX	08-DEC-1999	(first entry)
DT		
XX		
XX	Human breast tumour-associated protein 32.	
DE		
XX		
DE	Expressed sequence tag; EST; human; breast; cancer; cytostatic;	
KW	medicaments; gene therapy; treatment; fat metabolism.	
XX		
KW		
XX	Homo sapiens.	
OS		
XX		
XX	DE19813835-A1.	
PN		
XX		
PD	23-SEP-1999.	
XX		
XX	20-MAR-1998; 98DE-1013835.	
PF		
XX		
XX	20-MAR-1998; 98DE-1013835.	
PR		
XX		
XX	(META-) METAGEN GES GENOMFORSCHUNG MBH.	
PA		
XX		
PI	Specht T, Hinzmann B, Schmitt A, Pilarczyk C, Dahl E, Rosenthal A;	
PI	WPI; 1999-528979/45.	
DR	N-PSDB; AA233566.	
DR		
DR		
XX	Human nucleic acid sequences and protein products from normal breast	
PT	tissue, useful for breast cancer therapy	
PT		
XX	Claim 28; 173; 206pp; German.	
XX		
XX	This invention describes novel human nucleic acid sequences from normal	
XX	breast tissue which have cytostatic activity. The nucleic acid sequences	
CC	can be used to produce and isolate full-length gene sequences. They can	
CC	be used to express proteins, which can be used as tools to find an	
CC	activity against breast cancer. The sequences can be used in sense or	
CC	antisense form. They are especially useful for medicaments for gene	
CC	therapy to treat breast cancer and for treating illnesses associated	
CC	with fat metabolism. AA48456-Y48539 represent protein fragments encoded	
CC	by the expressed sequence tags described in the method of the invention.	
XX		
XX	Sequence 471 AA;	
SQ		

Query Match 62.6%; Score 2458; DB 20; Length 471;
Best Local Similarity 99.8%; Pred. No.5.3e-184;
Matches 470; Conservative 0; Mismatches 1; Indels 0; Gaps 0

Qy 289 NELKASGGEIKIHKMEOKENVPPGPVCITHQEGEKISANENSLAVRSTPAEDDSDSQV 348
| | | | |
Db 1 NELKASGGEIKIHKMEOKENVPPGPVCITHQEGEKISANENSLAVRSTPAEDDSDSQV 60
| | | | |

Qy	349	KSEVQOPVHPKPLSPDRASSLESSPPKMKFQAPARETCVECKTYYPMERLLANQQ	408
Db	61	KSEVQOPVHPKPLSPDRASSLESSPPKMKFQAPARETCVECKTYYPMERLLANQQ	120

Qy	409	VFHISFRCSYCNKLSLGTASYASLHGRIYCKPHEFNQLFKSGNYDEGFGHPRHDKDLWASK	468
Db	121	VFHISFRCSYCNKLSLGTASYASLHGRIYCKPHEFNQLFKSGNYDEGFGHPRHDKDLWASK	180

QY 469 NENEILTERPAQLANARETPHSPGVEDAPIAKGGVLAASMEAKASSQKEKDKPAETKKL 528
181 NENEILTERPAQLANARETPHSPGVEDAPIAKGGVLAASMEAKASSQKEKDKPAETKKL 240

Qy	^{16O}	529	RIAWPPTELGSSGSALEEGIKMSKPKWPPEDEISKEVPEDVDLDLKLRRSSLKERS	588
<hr/>				
Db		241	RIAWPPTELGSSGSALEEGIKMSKPKWPPEDEISKEVPEDVDLDLKLPRSSLKERS	300

[illegible]

Qy	649	GKTTWQNKESGKTGRSKRSGHSLNENLVNGASDEDDNSFLKQOSPQPKSLNWS	708
Db	361	GKTTWQNKESGKTGRSKRSGHSLNENLVNGASDEDDNSFLKQOSPQPKSLNWS	420

100

QY 709 SFVDNTFAEFTTQKQSDVELWEGEVVKLSVEEQIKRNRYDEDEDE 759
AAB95477
DB 421 SFVDNTFAEFTTQKQSDVELWEGEVVKLSVEEQIKRNRYDEDEDE 471

RESULT 8
ID AAB95477 standard; Protein; 457 AA.
AC AAB95477;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17992.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17992; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 457 AA;

Query Match 61.0%; Score 2394; DB 22; Length 457;
Best Local Similarity 100.0%; Pred. No. 5.3e-179;

	Matches	457; Conservative	0; Mismatches	0; Indels	0; Gaps	0;						
Qy	303	MEQKENVPPG	EVCI	THOEGEKI	SANENS	LA	VR	STPAEDD	SRD	SOVKSEVOQPVHPKPLS	362	
Db	1	MEQKENVPPG	EVCI	THOEGEKI	SANENS	LA	VR	STPAEDD	SRD	SOVKSEVOQPVHPKPLS	60	
Qy	363	PDSRASSLS	ESSPP	KAMK	FFQ	AP	ARE	TCV	ECQ	KTVYPMERLLANQQVPHI	SCFRCSYCNN	422
Db	61	PDSRASSLS	ESSPP	KAMK	FFQ	AP	ARE	TCV	ECQ	KTVYPMERLLANQQVPHI	SCFRCSYCNN	120
Qy	423	KLSLGT	YASL	HGR	IY	CK	PH	NQ	LFK	SKGNYDEGFGHPRHKLWASKNENE	EILRPAOLA	482
Db	121	KLSLGT	YASL	HGR	IY	CK	PH	NQ	LFK	SKGNYDEGFGHPRHKLWASKNENE	EILRPAOLA	180
Qy	483	NARE	TPH	SPG	VED	AP	IA	KV	GV	LAASMEAKASSQOEKEDKPAETKKLR	1AWPPPTLGGSSG	542
Db	181	NARE	TPH	SPG	VED	AP	IA	KV	GV	LAASMEAKASSQOEKEDKPAETKKLR	1AWPPPTLGGSSG	240
Qy	543	SALE	EGIK	MSK	PK	WPP	DE	DE	IS	KEP	VEDVLDLKKLRSSSLKERSRPPTVAASFQSTSV	602
Db	241	SALE	EGIK	MSK	PK	WPP	DE	DE	IS	KEP	VEDVLDLKKLRSSSLKERSRPPTVAASFQSTSV	300
Qy	603	KGP	KT	VS	PP	IR	KG	MS	MS	Q	SEESVGRVAERKQVENAKASKKNGVGT1TWONKESKGET	662
Db	301	KSP	KT	VS	PP	IR	KG	MS	MS	Q	SEESVGRVAERKQVENAKASKKNGVGT1TWONKESKGET	360
Qy	663	GKRS	KE	GH	SL	EM	EN	LV	ENG	ADSD	DDNSFLKQSPQEPKSLWSSVVDNTFAEFTTQ	722
Db	361	GKRS	KE	GH	SL	EM	EN	LV	ENG	ADSD	DDNSFLKQSPQEPKSLWSSVVDNTFAEFTTQ	420
Qy	723	NQ	KSD	VEL	WEG	EV	V	KLS	VEE	QIKRNRY	DEDEDE	759
Db	421	NQ	KSD	VEL	WEG	EV	V	KLS	VEE	QIKRNRY	DEDEDE	457

RESULT 9
AAU10979
ID AAU10979 standard; Protein; 457 AA.
XX
AC AAU10979;
XX
DT 12-MAR-2002 (first entry)
XX
DE Human sterol regulatory element binding protein 3.
XX
KW Human; sterol regulatory element binding protein 3; HSREBP-3;
KW hypothalamus.
XX
OS Homo sapiens.
XX
PN CN1309182-A.
XX
PD 22-AUG-2001.
XX
PF 17-FEB-2000; 2000CN-0111698.
XX
PR 17-FEB-2000; 2000CN-0111698.
XX
PA (NATU-) NATURAL HUMAN GENOME NANFANG RES CENT.
XX
PI Li Y, Xu S, Ren S;
XX
DR WPI; 2002-011822/02.
DR N-PSDB; AAS18588.
XX
PT Cholesterol regulatory factor binding protein and its coding sequence -
XX
PS Claim 2; Page 14; 27pp; Chinese.
XX
CC The invention relates to a novel human sterol regulatory element binding
CC protein 3 (HSREBP-3) expressed in human normal hypothalamic tissue and
CC its coding sequence. Also described is the process for preparing the
CC protein and nucleic acid sequence, and the method for detecting HSREBP-3

CC nucleic acid sequence and polypeptides. The present sequence represents
CC the amino acid sequence of human HSREBP-3 as described in the invention.
XX
SQ Sequence 457 AA;

Query Match 60.6%; Score 2378; DB 23; Length 457;
Best Local Similarity 99.3%; Pred. No. 9, 5e-178;
Matches 454; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 303 MEQKENVPPGPEVCITHGEKISANENSLAVRSTPAEDDSQVSEVQVHPKPLS 362
Db 1 MEQKENVPPGPEVCITHGEKISANENSLAVRSTPAEDDSQVSEVQVHPKPLS 60
QY 363 PDSRASSLSSEPPKAMKFFQAPARETCVECKTYPMERLLANQQVPHISCFRCSCYNN 422
Db 61 PDSRASSLSSEPPKAMKFFQAPARETCVECKTYPMERLLANQQVPHISCFRCSCYNN 120
QY 423 KLSLGTYSASLHGRYICKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENELRPAQLA 482
Db 121 KLSLGTYSASLHGRYICKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENELRPAQLA 180
QY 483 NARETPHSPGVGEDAPIAKVGLAASMEAKASSQEKEDKPAETKKLRIAWPPPTLGGSSG 542
Db 181 NARETPHSPGVGEDAPIAKVGLAASMEAKASSQEKENRPAETKKLRIAWPPPTLGGSSG 240
QY 543 SALESGIKMSKPKWPPDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFQSTSV 602
Db 241 SALESGIKMSKPKWPPDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFQSTSV 300
QY 603 KSPKTVSPPIKKGMSQSSSESUGRVAERKQVENAKSKNGNVGKTTWQNKESKGET 662
Db 301 KSPKTVSPPIKKGMSQSSSESUGRVAERKQVENAKSKNGNVGKTTWQNKESKGET 360
QY 663 GKRSEKHSLEMENENLVENGADSDSDNSFLKQOSPOEPKSLNWSFVDNTFAEERTTQ 722
Db 361 GKRSEKHSLEMENENLVENGADSDSDNSFLKQOSPOEPKSLNWSFVDNTFAEERTTQ 420
QY 723 NQKSDVELWEVGVKLSVEEQIKRNYDEDEDEE 759
Db 421 NQKSDVELWEVGVKLSVEEQIKRNYDEDEDEE 457

RESULT 10
ABG08559
ID ABG08559 standard; Protein: 476 AA.
XX AC ABG08559;
XX DT 13-FEB-2002 (first entry)
XX Novel human diagnostic protein #8550.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS72746.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID NO 38918; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 476 AA;
SQ
Query Match 51.2%; Score 2011; DB 22; Length 476;
Best Local Similarity 90.4%; Pred. No. 6, 3e-149;
Matches 405; Conservative 7; Mismatches 28; Indels 8; Gaps 7;
QY 320 QEKEKISANENSLAVRSTPAEDDS-RDSQVKSEVQVHPKPLSPDSRASSLSSESPPKA 378
Db 29 QEKEKISANENSLAVRSTPAEDDSPGDSQVKSEVQVHPKPLSPDSRASSLSSESPPKA 88
QY 379 MKXFPAPARETCVECKTYPMERLLANQQVPHISCFRCSCYNNKLSLGTYSASLHGRYIC 438
Db 89 MKXFPAPARETCVECKTYPMERLLANQQVPHISCFRCSCYNNKLSLGTYSASLHGRYIC 148
QY 439 KPHFNQLFKSKGNYDEGFGHRPHKDL-WASKNENELRPAQLANARETPHSPGVEDAP 497
Db 149 KPHFNQLFKSKGNYDEGFGHRPHKDLIGQAKMKTEILERQQLANARETPHSPGVEDAP 208
QY 498 IAKVGVLAASMEAKASSQEKEDKPAETKKLRIAWPPPTLGGSSGSALEEGIKMSKPKWP 557
Db 209 IAKVGVLAASMEAKASSQEKEDKPAETKKLRIAWPPPTLGGSSGSALEEGIKMSKPKWP 268
QY 558 PEDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFQSTSVKSPKTVSPPIKKGWS 617
Db 269 PEDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFQSTSVKSPKTVSPPIKKGWS 328
QY 618 MSQSEESVGGRVAERKQVENAKSKNGNVGKTTWQNKESKGETGK-RSEKHSLEMEN 676
Db 329 MSQSEESVGGRVAERKQVENAKSKNGNVGKTTWQNKESKGETGKRXKGVHSELEMEN 388
QY 677 ENLVENGADSD-EDNSF--LKQOSPOEPKSLNWS-SFVDN-TFAEERTTQNKSDVEL 731
Db 389 ENLVENGADSD-EDNSF--LKQOSPOEPKSLNWS-SFVDN-TFAEERTTQNKSDVEL 448
QY 732 WEQEVVVKLSVEEQIKRNYDEDEDEE 759
Db 449 WEQEVVVKLSVEEQIKRNYDEDEDEE 476

RESULT 11
AAB34206
ID AAB34206 standard; Protein: 299 AA.
XX
XX AC AAB34206;

XX DT 26-JAN-2001 (first entry)
XX DE Human secreted protein sequence encoded by gene 41 SEQ ID NO:174.
XX KW Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
XX KW antineumatic; antiproliferative; cytostatic; cardiant; vasotropic;
XX KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
XX KW fungicide; ophthalmological; vulnerary; gene therapy; autoimmune disease;
XX KW hyperproliferative disorder; neoplasm; cancer; cardiovascular disorder;
XX KW cerebrovascular disorder; angiogenesis; nervous system disorder;
XX KW infection; ocular disorder; wound healing; skin aging; food additive;
XX KW preservative.
XX OS Homo sapiens.
XX PN WO200056755-A1.
XX PD 28-SEP-2000.
XX PF 16-MAR-2000; 2000WO-US06830.
XX PR 19-MAR-1999; 99US-0125361.
XX PR 10-DEC-1999; 99US-0169910.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM, Komatsoulis G;
XX DR WPI; 2000-587661/55.
XX PT New isolated nucleic acid molecules encoding 49 human secreted proteins
XX PT used for preventing, treating or ameliorating medical conditions, for
XX PT diagnosing pathological conditions or as food additives or
XX PT preservatives -
XX PS Disclosure; Page 412-413; 419pp; English.
XX CC The polynucleotide sequences given in AAC59449 to AAC59497 encode the
XX CC human secreted proteins given in AAB34092 to AAB34140. AAB34141 to
XX CC AAB34216 represent human secreted polypeptide sequences and proteins
XX CC homologous to them, which are given in the exemplification of the present
XX CC invention. Human secreted proteins have activities based on the tissue
XX CC and cells the genes are expressed in. Examples of activities include:
XX CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
XX CC cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;
XX CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX CC and vulnerary. The polynucleotides and polypeptides can be used to
XX CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX CC in diagnosing a pathological condition or susceptibility to a
XX CC pathological condition. Disorders which are diagnosed or treated include
XX CC autoimmune diseases, hyperproliferative disorders e.g. neoplasms or
XX CC cancer of the breast or liver, cardiovascular disorders, cerebrovascular
XX CC disorders, angiogenesis, nervous system disorders, infections caused by
XX CC bacteria, viruses and fungi and ocular disorders. The polypeptides can
XX CC also be used to aid wound healing and epithelial cell proliferation, to
XX CC prevent skin aging due to sunburn, to maintain organs before
XX CC transplantation, for supporting cell culture of primary tissues, to
XX CC regenerate tissues and in chemotaxis. The polypeptides can also be used
XX CC as a food additive or preservative to increase or decrease storage
XX CC capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used
XX CC in the exemplification of the present invention.

QY 514 SQBQKEDKPAETKKLRIAWPPPTTELGGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDL 573
DB 61 SQBQKEDKPAETKKLRIAWPPPTTELGGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDL 120
QY 574 DLKLRSSSLKERSRPFVTAASQSTSVKSPKTVSPPIRKGSMSQSESVCGRAVER 633
DB 121 DLKLRSSSLKERSRPFVTAASQSTSVKSPKTVSPPIRKGSMSQSESVCGRAVER 180
QY 634 KOVENAKASKNGNVGKTTWQNKESKGETGRSKSGHSLMENENLVENGADSDSDNSF 693
DB 191 KOVENAKASKNGNVGKTTWQNKESKGETGRSKSGHSLMENENLVENGADSDSDNSF 240
QY 694 LKQSPQEPKSLNWSSFVDNTFAEEFTTQNKSOODVELWEGEVVVKLSVBEQIKRNRYY 752
DB 241 LKQSPQEPKSLNWSSFVDNTFAEEFTTQNKSOODVELWEGEVVVKLSVBEQIKRNRYY 299
RESULT 12
AAB56420
ID ABB56420 standard; Protein; 301 AA.
XX AC ABB56420;
XX DT 21-FEB-2002 (first entry)
XX DE Human cancer suppressor protein PP624.
XX KW Human; cancer suppressor; disease; cancer.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 21 /label= unknown
FT Misc-difference 40 /note= "Encoded by RCA"
FT Misc-difference 40 /label= unknown
FT Misc-difference 43 /note= "Encoded by TRC"
FT Misc-difference 43 /label= unknown
FT Misc-difference 43 /note= "Encoded by AKC"
CN1313297-A.
PD 19-SEP-2001.
PF 09-MAR-2000; 2000CN-0111948.
PR 09-MAR-2000; 2000CN-0111948.
PA (SHAN-) SHANGHAI INST ONCOLOGY.
XX Gu J, Yang S;
XX WPI; 2002-042185/06.
DR N-PSDB; ABI98973.
XX Human protein able to suppress growth of cancer cells and its coding
XX sequence -
XX PS Claim 1; Page 20 Disclosure; 37pp; Chinese.
XX CC The invention relates to novel human proteins (ABB56417-ABB56425) with
XX CC cancer suppressing function, the encoding polynucleotides
XX CC (AB198970-AB198978), the process for preparing the polypeptide, the
XX CC application of the polypeptide in treating diseases such as cancer, the
XX CC antagonist of the polypeptide and its medical function and the
XX CC application of the polynucleotide.
SQ Sequence 301 AA;

Query Match 37.5%; Score 1473; DB 23; Length 301;
Best Local Similarity 96.6%; Pred. No. 5e-107; 7; Indels 0; Gaps 0;
Matches 280; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 379 MKFQAPARETCVECKTQVYMERLLANQVPHISCFRCYCNKLSLGTYSASLHGRIVC 438
Db 1 MKFQAPARETCVECKTQVYMERLLANQVPHISCFRCYCNKLSLGTYSASLHGRIVC 60

QY 439 KPHFNLFKSGNYDEGFGHRRPKDQWASKNENEELERPAQLANARETPHSGVEDAPI 498
Db 61 KPHFNLFKSGNYDEGFGHRRPKDQWASKNENEELERPAQLANARETPHSGVEDAPI 120

QY 499 AKGVVLAASMEAKASSQOEKEDKPAETKRLRIAWPPPTLGGSSGSALEBEGIKMSKPKWPP 558
Db 121 AKGVVLAASMEAKASSQOEKEDKPAETKRLRIAWPPPTLGGSSGSALEBEGIKMSKPKWPP 180

QY 559 EDEISKPEVPEDVDLDLKKLRSSSLKRSRPTTVAASFQSTSVKSPKTVSPPIKQWSM 618
Db 181 EDEISKPEVPEDVDLDLKKLRSSSLKRSRPTTVAASFQSTSVKSPKTVSPPIKQWSM 240

QY 619 SQSEESVGRVAERKQVENAKSKKNGNVGKTTWQNKESKGETKRSKE 668
Db 241 SQSEESVGRVAERKQVENAKSKKNGNVGKTTWQNKESKGETKRSKE 290

RESULT 13
AAB54159
ID AAB54159 standard; Protein; 243 AA.
AC AAB54159;
DT 09-MAR-2001 (first entry)
XX Human pancreatic cancer antigen protein sequence SEQ ID NO:611.
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
OS Homo sapiens.
XX
XX WO200055320-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05989.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI; 2000-579444/54.
XX N-PSDB; AAC98924.
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
XX Claim 11; Page 1050-1051; 1379pp; English.
XX
XX AAC98773 to AAC9231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays

CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC9232 to AAC9240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 243 AA;
Query Match 31.9%; Score 1254; DB 21; Length 243;
Best Local Similarity 98.8%; Pred. No. 5.5e-90;
Matches 238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 321 EGEKISANENSLAVRSTPAEDDSRDSDQVKSEVQVHPKPLSPDRASSLSSESPPKAMK 380
Db 1 EGEKISANENSLAVRSTPAEDDSRDSDQVKSEVQVHPKPLSPDRASSLSSESPPKAMK 60

QY 381 KFOAPARETCVECKTQVYMERLLANQVPHISCFRCYCNKLSLGTYSASLHGRIVCKP 440
Db 61 KFOAPARETCVECKTQVYMERLLANQVPHISCFRCYCNKLSLGTYSASLHGRIVCKP 120

QY 441 HFNQLFKSGNYDEGFGHRRPKDQWASKNENEELERPAQLANARETPHSGVEDAPIAK 500
Db 121 HFNQLFKSGNYDEGFGHRRPKDQWASKNENEELERPAQLANARETPHSGVEDAPIAK 180

QY 501 VGVLAASMEAKASSQOEKEDKPAETKRLRIAWPPPTLGGSSGSALEBEGIKMSKPKWPPED 560
Db 181 VGVLAASMEAKASSQOEKEDKPAETKRLRIAWPPPTLGGSSGSALEBEGIKMSKPKWPPED 560

QY 561 E 561
Db 241 E 241

RESULT 14
AAB51883
ID AAB51883 standard; Protein; 232 AA.
AC AAB51883;
XX
XX 16-FEB-2001 (first entry)
XX
XX Human secreted protein sequence encoded by gene 5 SEQ ID NO:116.
XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neoplastic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; vulnery; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorders; cancer; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
KW Alzheimer's disease; ocular disorder; wound healing; skin aging.
OS Homo sapiens.
XX
XX WO2000061626-A1.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09066.
XX
XX 09-APR-1999; 99US-0128698.
XX 20-JAN-2000; 2000US-0176926.
XX
XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.
XX Rosen CA, Ruben SM, Komatsoulis G;
PI WPI: 2000-619227/59.
XX N-PSDB: AAC93483.
DR New nucleic acid molecules encoding 49 human secreted proteins for
XX diagnosing, preventing or ameliorating medical conditions and used for
PT food additives or preservatives -
XX
XX Disclosure: Page 22; 516pp; English.
XX
XX Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding
XX human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -
XX AAB51927 represent alternative polypeptides encoded by the genes, and
XX amino acid sequences with which they share homology. The genes and
XX proteins have activities dependent on the tissues and cells in which they
XX are expressed. Examples of their activities include immunosuppressive;
XX antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
XX vasotrophic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
XX virucide; fungicide; ophthalmological; and vulnerary. The secreted
XX proteins, polynucleotides, antagonists and agonists may be useful in
XX treating, preventing and/or diagnosing diseases and disorders such as
XX autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX angiodysplasia, nervous system disorders e.g. Alzheimer's disease,
XX infections caused by bacteria, viruses and fungi and ocular disorders
XX e.g. corneal infection. The polypeptides can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities, fat content, lipid, protein,
XX carbohydrate, vitamins, minerals, cofactors and other nutritional
XX components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are
XX used in the isolation and characterisation of the proteins and
XX polynucleotides of the invention.
XX
SQ Sequence 232 AA;
Query Match 29.5%; Score 1158; DB 21; Length 232;
Best Local Similarity 100.0%; Pred. No. 1.7e-82;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 473 EILERRPAQLANARETHPSGVEDAPITAKGVLAASMEAKASSQOEKEDPAETKKLRIAM 532
Db 1 EILERRPAQLANARETHPSGVEDAPITAKGVLAASMEAKASSQOEKEDPAETKKLRIAM 60
QY 533 PPTELSSGSSALEEGIKMSKPKWPPDEISKPEVPEVDLDDKLLRSSSLKERSRPT 592
Db 61 PPTELSSGSSALEEGIKMSKPKWPPDEISKPEVPEVDLDDKLLRSSSLKERSRPT 120
QY 593 VAASFQSTSVKSPKTSVPPPIRKQMSNSEQSEESVGGVRAERKQVENAKASKXNGNVGKTT 652
Db 121 VAASFQSTSVKSPKTSVPPPIRKQMSNSEQSEESVGGVRAERKQVENAKASKXNGNVGKTT 180
QY 653 WQNKESKGETGKRSKEGHSLEMEENLVENGADSDDEDDNSFLKQSQP 699
Db 181 WQNKESKGETGKRSKEGHSLEMEENLVENGADSDDEDDNSFLKQSQP 227

KW Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotrophic;
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulnerary; gene therapy; autoimmune disease;
KW hyperproliferative disorder; neoplasm; cancer; cardiovascular disorder;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW infection; ocular disorder; wound healing; skin aging; food additive;
KW preservative.
XX Cricetulus griseus.
OS WPI: 2000-587661/55.
XX
XX 28-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-US06830.
XX
XX 19-MAR-1999; 99US-0125361.
XX 10-DEC-1999; 99US-0169910.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI: 2000-587661/55.
XX
XX New isolated nucleic acid molecules encoding 49 human secreted proteins
XX used for preventing, treating or ameliorating medical conditions, for
XX diagnosing pathological conditions or as food additives or
XX preservatives -
XX
XX Disclosure: Page 411-412; 419pp; English.
XX
XX The polynucleotide sequences given in AAC59449 to AAC59497 encode the
XX human secreted proteins given in AAB34092 to AAB34140. AAB34141 to
XX AAB34216 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are given in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissue
XX and cells the genes are expressed in. Examples of activities include:
XX antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
XX cytostatic; cardiant; vasotrophic; cerebroprotective; neurotropic;
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX and vulnerary. The polynucleotides and polypeptides can be used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX in diagnosing a pathological condition or susceptibility to a
XX pathological condition. Disorders which are diagnosed or treated include
XX autoimmune diseases, hyperproliferative disorders e.g. neoplasms or
XX cancer of the breast or liver, cardiovascular disorders, cerebrovascular
XX disorders, angiogenesis, nervous system disorders, infections caused by
XX bacteria, viruses and fungi and ocular disorders. The polypeptides can
XX also be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The polypeptides can also be used
XX as a food additive or preservative to increase or decrease storage
XX capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used
XX in the exemplification of the present invention.
XX
SQ Sequence 299 AA;
Query Match 27.2%; Score 1068; DB 21; Length 299;
Best Local Similarity 71.0%; Pred. No. 3e-75;
Matches 213; Conservative 30; Mismatches 55; Indels 2; Gaps 2;
QY 454 EGFGRPHKDLWASKNENEILERPAPLANARETHPSGVEDAPITAKGVLAASMEAKAS 513
Db 1 EGFGRPHKDLWASKNENEILERPAPLANARETHPSGVEDAPITAKGVLAASMEAKAS 60
QY 514 SQQEKEDKPAETKKLRIAMPPPTTELSSGSSALEEGIKMSKPKWPPDEISKPEVPEVDL 573
Db 61 PQREKEDKPAETKKLRIAMPPAPLPPAELOSGSGTMEEGFKGSKPKPPPEDETSKLTALGVDL 120

QY	574	DLKLRSSSLKERSRPF	TVAA	SFQSTSVKSPKTVSPPIRKGWMSKSEQSEESVUGRVAER	633
Db	121	DLKLRSSSLKERSRPF	TVAA	SFRTSSLKSPKTLSPIRKGWMSKSEQNEESGGIVTAG	179
QY	634	KQVENAKASKKNGN	VGKTTWQNKES	-KGETGKRSKEGHSLEWENENLVENGADSDDEDDNS	692
Db	180	KQTENAKAPCKEN	VKSRWQREDSAREE	VARRSKDIHGVELESENFIENGANIYEGDSC	239
QY	693	FLKQSPQBP	KSUNWSSFVDNTFAEFTTQ	NKSQDVELWEGEVVKELSVVEEQIKENRY	752
Db	240	DLVQSP	PLDPKSPWSGFDTTPTTEFTTQ	NKSHDVGFREGEVVELSVVEEQIKENRY	299

Search completed: January 6, 2004, 09:47:18
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 6, 2004, 13:51:06 ; Search time 13109 Seconds
(without alignments)
11562.298 Million cell updates/sec
Title: US-09-890-549-16
Perfect score: 3705
Sequence: 1 ggcaccagagcagtagtg.....atttgaaaaaaaaaaaaa 3705
Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0
Searched: 2888711 seqs, 2045481386 residues
Word size : 0
Total number of hits satisfying chosen parameters: 5777422
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3611	97.5	3664	6	AX086386 Sequence
2	3611	97.5	3664	9	AF198454 Homo sapi
3	3548	95.8	3655	9	AF198454 Homo sapi
4	3114	84.0	3277	9	AK000335 Homo sapi
5	3098	83.6	3550	9	AF198455 Homo sapi
6	2647	71.4	2783	6	AX017487 Sequence
7	2647	71.4	2783	6	BD135177 Human nuc
8	2386	64.4	2536	9	AK000372 Homo sapi
9	2324	62.7	3348	9	AF157325 Homo sapi
10	2234	60.3	115345	9	AF157325 Homo sapi
11	1967	53.1	2121	9	BC010664 Homo sapi
12	1903	51.4	2667	9	AK000057 Homo sapi
13	1901	51.3	2164	9	BC001247 Homo sapi
14	1519	41.0	2207	6	BD160117 Primer fo
15	1519	41.0	2207	6	AK023649 Homo sapi
16	1462	39.5	2793	9	AK096172 Homo sapi
17	1296	35.0	1754	9	AF218025 Homo sapi
18	1109	29.9	28186	9	AC139016 Homo sapi
19	945	25.5	1021	6	AX012180 Sequence
20	714	19.3	777	11	G55637 SHGC-102035
21	604	16.3	641	6	AX011639 Sequence
22	390	10.5	390	6	BD113230 EST and e
23	343	9.3	343	6	AX378979 Sequence
24	343	9.3	343	6	AX379294 Sequence
25	325	8.8	389	6	BD025824 Sequence
26	276	7.4	279	6	AR269930 Sequence
27	248	6.7	413	6	AR270117 Sequence
28	238	6.4	250	11	GI1512 human STS S
29	236	6.4	532	6	BD154971 Primer fo
30	212	5.7	247	6	AX184947 Sequence
31	209	5.6	231	6	AX187815 Sequence
32	209	5.6	255	6	AX186750 Sequence
33	208	5.6	265	6	AX098162 Sequence
34	175	4.7	178	6	AX341839 Sequence
35	143	3.9	296	6	BD027230 Sequence
36	119	3.2	238	6	AR269618 Sequence
37	116	3.1	1109	9	AF245391S1 Homo sapi
38	87	2.3	213	6	AX198679 Sequence
39	86	2.3	86	6	AX209212 Sequence
40	42	1.1	840	6	AX323411 Sequence
41	35	1.0	425	10	CG922818 Cloning
42	31	0.8	2370	10	AF307845 Mus muscu
43	31	0.8	3232	10	BC031490 Mus muscu
44	31	0.8	3997	10	AF307844 Mus muscu
45	31	0.8	195301	2	AC134548 Mus muscu

ALIGNMENTS

RESULT 1.
AX086386
LOCUS AX086386 3664 bp DNA linear PAT 09-MAR-2001
DEFINITION Sequence 338 from Patent WO0112659.
ACCESSION AX086386
VERSION AX086386.1 GI:13275951
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Wiemann, S.
TITLE Human dna sequences
JOURNAL Patent: WO 0112659-A 338 22-FEB-2001;
Pred. No. is the number of results predicted by chance to have a

German Human Genome Project (DE)
Location/Qualifiers
source 1.3664
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606" 911 t
BASE COUNT 1210 a 734 c 809 g
ORIGIN
Query Match 97.5%; Score 3611; DB 6; Length 3664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3661; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 44 GCGTAGGTAGAGCGCGGACCTGTACAGGGCTGTAGCAGCGAGGAAAGCGGC 103
Db 1 GCGTAGGTAGAGCGCGGACCTGTACAGGGCTGTAGCAGCGAGGAAAGCGGC 60
Qy 104 TTTTAGCCAGTATTTCAAGTGTCTGTAGACAAGATGAATCATCTCCATTAATAGCGG 163
Db 61 TTTTAGCCAGTATTTCAAGTGTCTGTAGACAAGATGAATCATCTCCATTAATAGCGG 120
Qy 164 CAATGGACCTCACTATCATTTAGGGTTAACAGCCCAAGAACTTTCTTGTCAACAGAAC 223
Db 121 CAATGGACCTCACTATCATTTAGGGTTAACAGCCCAAGAACTTTCTTGTCAACAGAAC 180
Qy 224 AAGTCATCGGCTATTGTGGAATATTCTCCAAGTACCAGAAAGCAGCTGAAGAAACAAAC 283
Db 181 AAGTCATCGGCTATTGTGGAATATTCTCCAAGTACCAGAAAGCAGCTGAAGAAACAAAC 240
Qy 284 ATGGAGAAGAGAAAGTAAACCGGAAATCTCTCCAGACATTTAGAAAGGGGACCCCTG 343
Db 241 ATGGAGAAGAGAAAGTAAACCGGAAATCTCTCCAGACATTTAGAAAGGGGACCCCTG 300
Qy 344 ACTGTGTTAAGAGAGTGGGAGAACCCAGGCTGGGACAGAGTCTCACAGACTCT 403
Db 301 ACTGTGTTAAGAGAGTGGGAGAACCCAGGCTGGGACAGAGTCTCACAGACTCT 360
Qy 404 CTACGGACAGCAGCAGTGTAGATTAGCAGCAGAGCAGACCATCTCTCTGCTGAAGTGA 463
Db 361 CTACGGACAGCAGCAGTGTAGATTAGCAGCAGAGCAGACCATCTCTCTGCTGAAGTGA 420
Qy 464 AGCCACGCTGCTCTGGAGCCAAAGCTGACCAAGAAACAAATCCACCCAGACTCTAG 523
Db 421 AGCCACGCTGCTCTGGAGCCAAAGCTGACCAAGAAACAAATCCACCCAGACTCTAG 480
Qy 524 CTCAGGTCACTCTGAAGCCCTGTTAGGTCGATATCCACATCAAGACGCTGAG 583
Db 481 CTCAGGTCACTCTGAAGCCCTGTTAGGTCGATATCCACATCAAGACGCTGAG 540
Qy 584 GATCTTAAAGACCACTCAACAGAAAGTAAAGAAATGGAATTTGCTAGGAGATCCAGG 643
Db 541 GATCTTAAAGACCACTCAACAGAAAGTAAAGAAATGGAATTTGCTAGGAGATCCAGG 600
Qy 644 CATGAAGTAGAAAAATCAGAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAA 703
Db 601 CATGAAGTAGAAAAATCAGAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAA 660
Qy 704 TATTAATGTCGCTGAACAGCCTTAAGATGATGTTTGAAGAAAGTGAAACCAACTCAAACT 763
Db 661 TATTAATGTCGCTGAACAGCCTTAAGATGATGTTTGAAGAAAGTGAAACCAACTCAAACT 720
Qy 764 AAGATTCTCGGGCCCAAGCCGAGTGAAGTGAAGGAGATCTCTGAAACAGCTAT 823
Db 721 AAGATTCTCGGGCCCAAGCCGAGTGAAGTGAAGGAGATCTCTGAAACAGCTAT 780
Qy 824 TCTCTAGATGACCTGGAAATAGGCCAGGTCAGTTGTCATCTTCTACATTTGACTCGGAG 883
Db 781 TCTCTAGATGACCTGGAAATAGGCCAGGTCAGTTGTCATCTTCTACATTTGACTCGGAG 840
Qy 884 AAAAATGAGAGTAGCGAAATCTGGAACTTCCAGCTCTCAGAAACCTCTATAAAGGAT 943
Db 841 AAAAATGAGAGTAGCGAAATCTGGAACTTCCAGCTCTCAGAAACCTCTATAAAGGAT 900

Qy 944 CGAATGGCCAAAGTACCAGGACAGCTGTGTCCAAAACAAAGCAGCTCAACCAACTATACAAT 1003
Db 901 CGAATGGCCAAAGTACCAGGACAGCTGTGTCCAAAACAAAGCAGCTCAACCAACTATACAAT 960
Qy 1004 GAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGAGATGTG 1063
Db 961 GAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGAGATGTG 1020
Qy 1064 CCCCAGGTCCTGAGGTCCTGATCACCCATCAGAAAGGGGAAAGATTTCTGCATAATGAG 1123
Db 1021 CCCCAGGTCCTGAGGTCCTGATCACCCATCAGAAAGGGGAAAGATTTCTGCATAATGAG 1080
Qy 1124 AATAGCCTGGCAGTCCGTTCCACCCCTGCGAAGATGACTCCCGTGAATCCAGGTTAAG 1183
Db 1081 AATAGCCTGGCAGTCCGTTCCACCCCTGCGAAGATGACTCCCGTGAATCCAGGTTAAG 1140
Qy 1184 AGTAGGTTCAACAGCCTGTCTATCCCAAGCCTAAGTCCAGATTCAGAGCCTCCAGT 1243
Db 1141 AGTAGGTTCAACAGCCTGTCTATCCCAAGCCTAAGTCCAGATTCAGAGCCTCCAGT 1200
Qy 1244 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGATTTTCAGGCACCTGCAAGAGAGACC 1303
Db 1201 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGATTTTCAGGCACCTGCAAGAGAGACC 1260
Qy 1304 TGCCTGGAATGTCAGAGACAGTCTATCCAATGAGAGCTCTCTTGGCCAAACAGCAGGTG 1363
Db 1261 TGCCTGGAATGTCAGAGACAGTCTATCCAATGAGAGCTCTCTTGGCCAAACAGCAGGTG 1320
Qy 1364 TTTACATCAGCTGCTCCGTTGCTCTATTGCAACAAACAACTCAGTCTAGGAACATAT 1423
Db 1321 TTTACATCAGCTGCTCCGTTGCTCTATTGCAACAAACAACTCAGTCTAGGAACATAT 1380
Qy 1424 GCATCTTTACATGGAAGATCTATTGTAAGCTCACTTCAATCAACTCTTTAAATCTAAG 1483
Db 1381 GCATCTTTACATGGAAGATCTATTGTAAGCTCACTTCAATCAACTCTTTAAATCTAAG 1440
Qy 1484 GCAACTATTGATGAAGGCTTTGGGCACAGACACACAAGGATCTATGGCAAGCAAAAT 1543
Db 1441 GCAACTATTGATGAAGGCTTTGGGCACAGACACACAAGGATCTATGGCAAGCAAAAT 1500
Qy 1544 GAAAAAGAGAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGGAGACCCCTCAC 1603
Db 1501 GAAAAAGAGAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGGAGACCCCTCAC 1560
Qy 1604 AGCCAGGGGTAGAGATGCCCCCTATTGCTAAGTGGGTGCTCTGGTGCAGATGAGAA 1663
Db 1561 AGCCAGGGGTAGAGATGCCCCCTATTGCTAAGTGGGTGCTCTGGTGCAGATGAGAA 1620
Qy 1664 GCAAGGCTCTCTCTCAGCAGAGAGAGAGACAGCCAGCTGAACCAAGAGCTGAGG 1723
Db 1621 GCAAGGCTCTCTCTCAGCAGAGAGAGAGAGACAGCCAGCTGAACCAAGAGCTGAGG 1680
Qy 1724 ATCGCTGGCCACCCCTCACTGAAGTTCAGGAAGTTCAGGAAGTTCAGGAAGAGGATC 1783
Db 1681 ATCGCTGGCCACCCCTCACTGAAGTTCAGGAAGTTCAGGAAGTTCAGGAAGAGGATC 1740
Qy 1784 AAAATGTCAAAGCCCAATGGCTCTCTGAAGCAGAAATCAGCAAGCCCAAGTTCCTGAG 1843
Db 1741 AAAATGTCAAAGCCCAATGGCTCTCTGAAGCAGAAATCAGCAAGCCCAAGTTCCTGAG 1800
Qy 1844 GATGTCGATCTAGATCTGAAGAGCTTAAGAGATCTTCTTCACTGAAGGAAAGAGCCGC 1903
Db 1801 GATGTCGATCTAGATCTGAAGAGCTTAAGAGATCTTCTTCACTGAAGGAAAGAGCCGC 1860
Qy 1904 CCATTTCATGTAGAGCTTCATTTCAAAGCCTCTGTCAGAGCCCAAAACCTGTGTCC 1963
Db 1861 CCATTTCATGTAGAGCTTCATTTCAAAGCCTCTGTCAGAGCCCAAAACCTGTGTCC 1920
Qy 1964 CCACCTATCAGGAAAGGCTGGAGCATGTGAGAGAGTGAAGAGTCTGTGGTGGGAGA 2023
Db 1921 CCACCTATCAGGAAAGGCTGGAGCATGTGAGAGAGTGAAGAGTCTGTGGTGGGAGA 1980
Qy 2024 GTTGCAAGAAAGAAACAAATGGAAGTTCCTTAAGAGAAATGCGGAAATGTGGGA 2083


```

/db_xref="taxon:9606"
/map="12"
/clone="DKFZp586i1918"
/tissue_type="uterus"
/clone_lib="586 (synonym: hute1). Vector pSport1; host
DH10B; sites NotI + SalI/MluI"
/dev_stage="adult"
1..3664
/gene="DKFZp586i1918"
94..2373
/gene="DKFZp586i1918"
/note="Similarity to transcription factor SF3"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB66845.1"
/db_xref="GI:12053117"
/db_xref="SPTREMBL:Q9UHB6"
/translation="MESSFNRKQTSLRVTAKELSLVNKNKSAIVEIFSKYQKA
AETNMKRSNTENLSQHRKGLTIVLKKWENPGLAESHTLSRNSSTEIRHAD
HPFAEVTSHAAGKADDEQIHPRSRLRSPPEALVQGRPHIKDGEDLKHSTESKK
MENCLGESHEVKESEISENTDASGKIEKYNVPLNRLNMPEKGEPTOTKILRAQSR
ASGRKISENSYSLDLEIGPGOLSSSTFDSKVESRRLNLELPLRSETSIKDRMAKYOA
AYSKOSSSNYNYNELKASGGEIKIHMEQKENVPPPEVCITHOGEKISANESLAV
RSTPAEDDSRDSQVEVOQPPVHPPLSPDSRASLSSESPKMKKFOARETCE
CQKTYPMERLLANQQVPHISCFRSCYCNKLSLGTYSALHGRYICKPFPNQLFKSG
NYDEGFHRPHDLWASKNENEIILERPQLANARETPHSPVEDAPIAKGVLSKSM
EAKASQOEKDEKPAETTKLRTAMPPELGLSSGALBEGIKMSKPKWPPDEISKPE
VPEDVDLKLKLRSSSLKERSRPTVAASFOSVSKPTVSPPIRKGMWSRQSEE
SVGGRVARKQVENAKAKNGNVGKTTWQNKESGETGKRSKEGHSLEMNENLVEN
GADSDDDNSFLKQSPQEPKSLNWSFVDNTFAEFTTQNKQSDVELMEGEVVKEL
SVEEQIKRNYRYYDEDEE"
3637..3642
/gene="DKFZp586i1918"
3647
/polyA_signal
/polyA_site
BASE COUNT 1210 a 734 c 809 g 911 t
ORIGIN
Query Match 97.5%; Score 3611; DB 9; Length 3664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3661; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 44 GCGTAGGTAGAGCGCGGACCTGTGACAGAGGTGTGTAGCAGCGCAGAGGAAAGCGCG 103
DB 1 GCGCTAGGTAGAGCGCGGACCTGTGACAGAGGTGTGTAGCAGCGCAGAGGAAAGCGCG 60
QY 104 TTTTACCGAGGTATTTTACGTGTCTGTAGACAGATGGAATCATCTTCCATTTAATAGACGG 163
DB 61 TTTTACCGAGGTATTTTACGTGTCTGTAGACAGATGGAATCATCTTCCATTTAATAGACGG 120
QY 164 CAATGGACCTCACTATCATTGAGGTAAACAGCCAAAGACATTTCTTGTCTCAACAGAAC 223
DB 121 CAATGGACCTCACTATCATTGAGGTAAACAGCCAAAGACATTTCTTGTCTCAACAGAAC 180
QY 224 AAGTCATCGGCTATTGTGGAATATTCTCCAAAGTACAGAAAGAGAGCTGAAGAAACAAAC 283
DB 181 AAGTCATCGGCTATTGTGGAATATTCTCCAAAGTACAGAAAGAGAGCTGAAGAAACAAAC 240
QY 284 ATGGAGAAGAGAGAGTAAACCCGAAATCTCTCCAGCATTTTAGAAGGGGACCCCTG 343
DB 241 ATGGAGAAGAGAGAGTAAACCCGAAATCTCTCCAGCATTTTAGAAGGGGACCCCTG 300
QY 344 ACTGTGTTAAAGNAGAGTGGGAGACCCAGGCTGGGAGAGTCTTCCACAGACTCT 403
DB 301 ACTGTGTTAAAGNAGAGTGGGAGACCCAGGCTGGGAGAGTCTTCCACAGACTCT 360
QY 404 CTACGGAACAGAGACTGAGATTAGGACAGAGACCATCTCTCTGCTGAAGTGACA 463
DB 361 CTACGGAACAGAGACTGAGATTAGGACAGAGACCATCTCTCTGCTGAAGTGACA 420
QY 464 AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGAACAAATCCACCCAGATCTAGA 523
DB 421 AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGAACAAATCCACCCAGATCTAGA 480

```

```

QY 524 CTGAGGTCACTCTCTGAAGCCCTCGTTCAAGGTCGATATATCCACATCAAGGACGGTGAG 583
DB 481 CTGAGGTCACTCTCTGAAGCCCTCGTTCAAGGTCGATATATCCACATCAAGGACGGTGAG 540
QY 584 GATCTTTAAAGACCACTCAACAGAAAGTAAATAAATGGAATAATGTTCTAGGAGAAATCCAGG 643
DB 541 GATCTTTAAAGACCACTCAACAGAAAGTAAATAAATGGAATAATGTTCTAGGAGAAATCCAGG 600
QY 644 CATGAAGTGAATAAATCAGAAATCAGTGAATAACACAGATGCTTTCGGGCAAAATAGAGAAA 703
DB 601 CATGAAGTGAATAAATCAGAAATCAGTGAATAACACAGATGCTTTCGGGCAAAATAGAGAAA 660
QY 704 TATAATGTTCCGGTGAACAGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACT 763
DB 661 TATAATGTTCCGGTGAACAGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACT 720
QY 764 AAGATTCTCCGGGCCCAAAAGCGAAGTGAAGTGAAGGAAGATCTCTGAAAAACAGCTAT 823
DB 721 AAGATTCTCCGGGCCCAAAAGCGAAGTGAAGTGAAGGAAGATCTCTGAAAAACAGCTAT 780
QY 824 TCTCTAGATGACCTGGAAATAGGCCCAAGGTCAGTTGTCATCTTCTACATTTGACTCGGAG 883
DB 781 TCTCTAGATGACCTGGAAATAGGCCCAAGGTCAGTTGTCATCTTCTACATTTGACTCGGAG 840
QY 884 AAAAAATCAGAGTAGACGAAATCTGGAACTTCCACGCTCTCAGAAAACTCTATAAAGGAT 943
DB 841 AAAAAATCAGAGTAGACGAAATCTGGAACTTCCACGCTCTCAGAAAACTCTATAAAGGAT 900
QY 944 CGAATGCCCAAGTACCAAGGAGCTGTGTCCAAACAAAGACAGCTCAACCAACTATACAAAT 1003
DB 901 CGAATGCCCAAGTACCAAGGAGCTGTGTCCAAACAAAGACAGCTCAACCAACTATACAAAT 960
QY 1004 GAGCTGAAAGCCAGTGTGTGGCGAAATCAAAATTCATAAATGGAGCAAAAGGAGAAATGTG 1063
DB 961 GAGCTGAAAGCCAGTGTGTGGCGAAATCAAAATTCATAAATGGAGCAAAAGGAGAAATGTG 1020
QY 1064 CCCCCAGGTCTCTGAGGTCTGCATCACCCCTGCGCAAGATGACTCCCGTGACTCCAGGTTAAG 1183
DB 1021 CCCCCAGGTCTCTGAGGTCTGCATCACCCCTGCGCAAGATGACTCCCGTGACTCCAGGTTAAG 1140
QY 1124 AATAGCCTGCGAGTCCCTTCCACCCCTGCGCAAGATGACTCCCGTGACTCCAGGTTAAG 1183
DB 1081 AATAGCCTGCGAGTCCCTTCCACCCCTGCGCAAGATGACTCCCGTGACTCCAGGTTAAG 1140
QY 1184 AGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCCAGT 1243
DB 1141 AGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCCAGT 1200
QY 1244 CTTTCTGAAAGTCTCTCTCCCAAGGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACC 1303
DB 1201 CTTTCTGAAAGTCTCTCTCCCAAGGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACC 1260
QY 1304 TGGCTGGAATGTGAGAAAGACAGTCTATCCCAATGGAGGCTCTTTGGCCCAACACGAGGTTG 1363
DB 1261 TGGCTGGAATGTGAGAAAGACAGTCTATCCCAATGGAGGCTCTTTGGCCCAACACGAGGTTG 1320
QY 1364 TTTTCAATCAGCTGCTTCCGTGCTCTCTATTGCAACCAACAACTCAGTCTAGGAAACATAT 1423
DB 1321 TTTTCAATCAGCTGCTTCCGTGCTCTCTATTGCAACCAACAACTCAGTCTAGGAAACATAT 1380
QY 1424 GCATCTTTTATGGAAGAATCTATTGTAAGCTTCACTTTCAATCAACTCTTTTAAATCTAAG 1483
DB 1381 GCATCTTTTATGGAAGAATCTATTGTAAGCTTCACTTTCAATCAACTCTTTTAAATCTAAG 1440
QY 1484 GGCACCTATGATGAAGCTTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAACAAAT 1543
DB 1441 GGCACCTATGATGAAGCTTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAACAAAT 1500
QY 1544 GAAAAAGAGAGATTTTGGAGAGACCCAGCCAGCTTGCAAATGCAAGGGAGAGCCCTCAC 1603
DB 1501 GAAAAAGAGATTTTGGAGAGACCCAGCCAGCTTGCAAATGCAAGGGAGAGCCCTCAC 1560

```

QY 1604 AGCCAGGGGTAGAGATGCCCTATTGCTAAGGTGGTGTCTCGCTGCAAGTATGAA 1663
DB 1561 AGCCAGGGGTAGAGATGCCCTATTGCTAAGGTGGTGTCTCGCTGCAAGTATGAA 1620
QY 1664 GCAAGGCCCTCTCTCAGCAGGAGAGAGAACAGAGCCAGCTGAAACCAAGAGCTGAGG 1723
DB 1621 GCAAGGCCCTCTCTCAGCAGGAGAGAGAACAGAGCCAGCTGAAACCAAGAGCTGAGG 1680
QY 1724 ATCGGCTGCCACCCCTCACTGAACTTGAAGTTCAGGAAGTGCCTTGGAGAGAGGATC 1783
DB 1681 ATCGGCTGCCACCCCTCACTGAACTTGAAGTTCAGGAAGTGCCTTGGAGAGAGGATC 1740
QY 1784 AAAATGTCAAGGCCCAATGGCTCTCTGAGAGAGAAATCAGCAAGCCCGAGATTCCTGAG 1843
DB 1741 AAAATGTCAAGGCCCAATGGCTCTCTGAGAGAGAAATCAGCAAGCCCGAGATTCCTGAG 1800
QY 1844 GATGTCGATCTAGATCTGAAGAAGCTTAAGACGATCTTCTTCACTGAAGGAAAGAGCCGC 1903
DB 1801 GATGTCGATCTAGATCTGAAGAAGCTTAAGACGATCTTCTTCACTGAAGGAAAGAGCCGC 1860
QY 1904 CAATTCACTGTAGAGCTTCAATTTCAAGAGCCTCTGTCAAGAGCCCAAAAACCTGTGTC 1963
DB 1861 CAATTCACTGTAGAGCTTCAATTTCAAGAGCCTCTGTCAAGAGCCCAAAAACCTGTGTC 1920
QY 1964 CCACCTATCAGGAAGGCTGGAGCATGTGAGAGAGAGTGAAGTCTGTGGTGGAGA 2023
DB 1921 CCACCTATCAGGAAGGCTGGAGCATGTGAGAGAGAGTGAAGTCTGTGGTGGAGA 1980
QY 2024 GTTGCAAGAAAGAAACAAAGTGGAAATGCCAAGGCTTCTTAAGAGAAATGGGAATGGGA 2083
DB 1981 GTTGCAAGAAAGAAACAAAGTGGAAATGCCAAGGCTTCTTAAGAGAAATGGGAATGGGA 2040
QY 2084 AAAACAACTGCGCAAAAACAAAGAAATCTAAGAGAGAGACAGGGAAGAGAAAGT 2143
DB 2041 AAAACAACTGCGCAAAAACAAAGAAATCTAAGAGAGAGACAGGGAAGAGAAAGT 2100
QY 2144 CATAGTTTGGAGATGGAGATGAGATCTTGTAGAAATGTGAGACTCCGATGAAGAT 2203
DB 2101 CATAGTTTGGAGATGGAGATGAGATCTTGTAGAAATGTGAGAAATGTGAGACTCCGATGAAGAT 2160
QY 2204 GATAACAGCTTCTCAAAACAAATCTCCACAAAGCCCAAGTCTCTGAATTTGCTCAGT 2263
DB 2161 GATAACAGCTTCTCAAAACAAATCTCCACAAAGCCCAAGTCTCTGAATTTGCTCAGT 2220
QY 2264 TTTGTAGACACACCTTTGCTGAAGAAATCTACTCAGAAATCCAGATCCAGAGTGTG 2323
DB 2221 TTTGTAGACACACCTTTGCTGAAGAAATCTACTCAGAAATCCAGATCCAGAGTGTG 2280
QY 2324 GAACCTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAAT 2383
DB 2281 GAACCTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAAT 2340
QY 2384 CGGTATTATGATGAGATGAGATGAGAGTGAACAAATGCAATGATGCTGGGCTTAAA 2443
DB 2341 CGGTATTATGATGAGATGAGATGAGAGTGAACAAATGCAATGATGCTGGGCTTAAA 2400
QY 2444 TTCATGTTAGTGTAGCGAGCCACTGCCCTTTGTCAAAATGTGATGACATGAAGCAGTA 2503
DB 2401 TTCATGTTAGTGTAGCGAGCCACTGCCCTTTGTCAAAATGTGATGACATGAAGCAGTA 2460
QY 2504 TCCAGCATGAAATGTAAATTTACTTGAAGTAACTTTGGAAAAGAAATTCCTTCTTAAAT 2563
DB 2461 TCCAGCATGAAATGTAAATTTACTTGAAGTAACTTTGGAAAAGAAATTCCTTCTTAAAT 2520
QY 2564 CAAAAACAAAAACAAAAACAAATCTAAATCTAGAGATAACTTTACTTA 2623
DB 2521 CAAAAACAAAAACAAAAACAAATCTAAATCTAGAGATAACTTTACTTA 2580
QY 2624 AATTCTTCATTTAGCAGTGTATATCATAGTGTGTAAGGCTTGTAACTGGGGAAA 2683
DB 2581 AATTCTTCATTTAGCAGTGTATATCATAGTGTGTAAGGCTTGTAACTGGGGAAA 2640
QY 2684 TATTCCACCTGATAATGCCAGATTTCTACTGTATTCCTCAAAAGGCAATATTAGGTAGA 2743

DB 2641 TATTCCACCTGATAATAGCCAGATTTCTACTGTATTTCCAAAAGGCAATATTAGGTAGA 2700
QY 2744 TAGATGATTAGTAGTATATTCTTACACACTATTATTGGAAATTAGAGAACATACAGAAGAA 2803
DB 2701 TAGATGATTAGTAGTATATTCTTACACACTATTATTGGAAATTAGAGAACATACAGAAGAA 2760
QY 2804 TTTAGGGCTTTAAACATTACGACTGAATGCACTTTAGTATATAAAGGGCACAGTTTGTATAT 2863
DB 2761 TTTAGGGCTTTAAACATTACGACTGAATGCACTTTAGTATATAAAGGGCACAGTTTGTATAT 2820
QY 2864 TTTTAAATGAATACCAATTTAAATTTTGTAGTATTACCTGTTAAGAGATTTATTAGTCTT 2923
DB 2821 TTTTAAATGAATACCAATTTAAATTTTGTAGTATTACCTGTTAAGAGATTTATTAGTCTT 2880
QY 2924 TAAATTTTGTAGTATTATTTCTGTGTATATATATAGAGAAATTTACTACTTTTATGTC 2983
DB 2881 TAAATTTTGTAGTATTATTTCTGTGTATATATATAGAGAAATTTACTACTTTTATGTC 2940
QY 2984 CTGCTCTCTAAACTACATCTGAACTCGAGCTCTGAGGTATATAACACAGAGCACTTT 3043
DB 2941 CTGCTCTCTAAACTACATCTGAACTCGAGCTCTGAGGTATATAACACAGAGCACTTT 3000
QY 3044 TTGAGGCAATTTGAAAACCAACCTACACTCTTCGGTGTCTTAGAGAGATCTGCTGTCTCC 3103
DB 3001 TTGAGGCAATTTGAAAACCAACCTACACTCTTCGGTGTCTTAGAGAGATCTGCTGTCTCC 3060
QY 3104 AAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTTCTTTTCT 3163
DB 3061 AAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTTCTTTTCT 3120
QY 3164 GGTGATATCTGCTCTCTATAATTTACTGAAAGCTGCAATATTATTTAGTAATACCTTCGGG 3223
DB 3121 GGTGATATCTGCTCTCTATAATTTACTGAAAGCTGCAATATTATTTAGTAATACCTTCGGG 3180
QY 3224 ATCACTGTCCCCCATCTTCGGTGTAGAGCAAGTGAAGAGTTTAAAGGAGGAGAGAA 3283
DB 3181 ATCACTGTCCCCCATCTTCGGTGTAGAGCAAGTGAAGAGTTTAAAGGAGGAGAGAA 3240
QY 3284 AGAAGTGTCTTACACACTGTGAGCTCAGACCTCTAAACCTGTATTTCCTTATGATGTC 3343
DB 3241 AGAAGTGTCTTACACACTGTGAGCTCAGACCTCTAAACCTGTATTTCCTTATGATGTC 3300
QY 3344 CCTTTTGGAGACACTAAATTTTAAATCTTACTAGCTCTGAAATATATTGATTTTATC 3403
DB 3301 CCTTTTGGAGACACTAAATTTTAAATCTTACTAGCTCTGAAATATATTGATTTTATC 3360
QY 3404 ACAGTATTTCTCAGGTTGAAATTTAAACCACTATAGGCTTTTCTTTGGGATGATTTCTA 3463
DB 3361 ACAGTATTTCTCAGGTTGAAATTTAAACCACTATAGGCTTTTCTTTGGGATGATTTCTA 3420
QY 3464 GTCTTAAGTTTGGGACATTTATAA CTTGAGTACATTTTGTGTACACAGTTGATATTC 3523
DB 3421 GTCTTAAGTTTGGGACATTTATAA CTTGAGTACATTTTGTGTACACAGTTGATATTC 3480
QY 3524 AAATTTGATGGATGGAGGAGAGGTGCTTAAGCTGTAGGCTTTTCTTTGACTGCAAT 3583
DB 3481 AAATTTGATGGATGGAGGAGAGGTGCTTAAGCTGTAGGCTTTTCTTTGACTGCAAT 3540
QY 3584 TATAGAGATTTAGCTTTAATATTTTGTAGAGATGTAACATCTGCTTTCTTAGTCTTA 3643
DB 3541 TATAGAGATTTAGCTTTAATATTTTGTAGAGATGTAACATCTGCTTTCTTAGTCTTA 3600
QY 3644 CCTAGTCTGAACATTTTATTCATATAAAGATTTTAAATTTAAATTTGAAAAA 3703
DB 3601 CCTAGTCTGAACATTTTATTCATATAAAGATTTTAAATTTAAATTTGAAAAA 3660
QY 3704 AA 3705
DB 3661 AA 3662

RESULT 3

AF198454
LOCUS AF198454 3655 bp mRNA linear PRI 10-JAN-2000
DEFINITION Homo sapiens epithelial protein lost in neoplasm beta (EPLIN) mRNA, complete cds.
ACCESSION AF198454
VERSION AF198454.1 GI:6685006
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3655)
Maul, R.S. and Chang, D.D.
EPLIN, epithelial protein lost in neoplasm
JOURNAL Oncogene 18 (54), 7838-7841 (1999)
MEDLINE 20087188
PUBMED 10618726
REFERENCE 2 (bases 1 to 3655)
Maul, R.S. and Chang, D.D.
Direct Submission
JOURNAL Submitted (25-OCT-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA
FEATURES
source
1..3655
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q13"
1..3655
/gene="EPLIN"
102..2381
/gene="EPLIN"
/codon_start=1
/product="epithelial protein lost in neoplasm beta"
/protein_id="AAP3755.1"
/db_xref="GI:6685007"
/translation="MESSPFRNQWTSLSLRVTAKELSLVKNKSSAIVEIFSKYKA
AERTNMEKKSNTENSLNKHGKGLTLVKKKNWENPGLGAEHSHTDSLNSSTIRHEAD
HPEAVENTSHAASGAKADQEEQIHPRSLRPPALVQGRYPHFKDGEDLKHSTESKK
MENCLGSRHEVEKSEISENTDASGKIERIENFLNRLKMFKEGPEPTQKILRAOSRS
ASRKKISENSYDLEIGQLSSSTFSEKVESRNLELPLRSETSIKDRMAYQA
AVKQSSNTYTHLKSAGGEIKIHKMEQENVPPEVCITHQEGEKISANESLAV
RSTPAEDSDSOVKEVQVPHKPLSPDRASSLSSESPPKAMKKFOAPARETCVE
COKTVYPMERLLANQOVPHISCFRCSVNNKLSLGTVASLHGRIYCKPHNQLFKSKG
NYDEGFGRPHKOLWASKNENELBERPAQLANARETPSPGVEDAPIAKVGLVLAASG
EKASSQEKEDPAETKURIAWPPPTTELGSSGSALEEGIKMSKPKWPEDEISKPE
VPEDVDLKLKRRSSSLKERSRPFVVAASFQTSVKSPKTVSPPIKMGMSSEQSEE
SVGRVAERQVENAKKNGVGTWQNKESKGETGKRSKEGHSLEMNENLVLEN
GADSDDDNSFLKQSPQEPKSLNWSFVNDNTFAEEFTTQNKQSDVELWEGEVKEL
SVDEQIKRNYDEDEDEE"
polyA_signal 3632..3637
/gene="EPLIN"
BASE COUNT 1197 a 737 c 810 g 911 t
ORIGIN
Query Match 95.8%; Score 3548; DB 9; Length 3655;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3598; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 92 AGGAAGCGCGCTTTTAGCCAGGTATTTACGTGCTGTAGACAAGATGGAATCATCTCCA 151
DB 57 AGGAAGCGCGCTTTTAGCCAGGTATTTACGTGCTGTAGACAAGATGGAATCATCTCCA 116
QY 152 TTAAATAGCGGCAATGGAGCTCCTATCATATGAGGGTAACACGCCAAAGAACTTCTCTT 211
DB 117 TTAAATAGCGGCAATGGAGCTCCTATCATATGAGGGTAACACGCCAAAGAACTTCTCTT 176
QY 212 GTCAACAGAACAGTATCGGCTATTGTGGAATATTTCCAAATACCGAAAGAGCT 271
DB 177 GTCAACAGAACAGTATCGGCTATTGTGGAATATTTCCAAATACCGAAAGAGCT 236
QY 272 GAAGAAACAAACATGGAGAGAGAGAGAGTAACACGCCAAATCTCTCCAGCAGCTTTAGA 331

237 GAAGAAACAAACATGGAGAGAGAGAGTAACACGCCAAATCTCTCCAGCAGCTTTAGA 296
QY 332 AAGGGGACCCCTGACTGTGTTAAAGAGAGAGTGGGAGAACCCAGGGCTGGGAGCAGTCT 391
DB 297 AAGGGGACCCCTGACTGTGTTAAAGAGAGAGTGGGAGAACCCAGGGCTGGGAGCAGTCT 356
QY 392 CACACAGAGCTCTCTACGGAACAGCAGCAGCTGAGATTAGGCACAGAGCAGACCATCTCTCT 451
DB 357 CACACAGAGCTCTCTACGGAACAGCAGCAGCTGAGATTAGGCACAGAGCAGACCATCTCTCT 416
QY 452 GCTGAAGTGAACAGCAGCAGCTGCTTCTGGAGCAAGCTGACCAAGAGAGAGCAAAATCCAC 511
DB 417 GCTGAAGTGAACAGCAGCAGCTGCTTCTGGAGCAAGCTGACCAAGAGAGAGCAAAATCCAC 476
QY 512 CCAGATCTAGACTCAGCTCAGCTCCTCTGAAGCCCTCGTTACGGGTGATATCCCCACATC 571
DB 477 CCAGATCTAGACTCAGCTCAGCTCCTCTGAAGCCCTCGTTACGGGTGATATCCCCACATC 536
QY 572 AAGGACGGTGAAGATCTTAAAGAGCAGCTCAACAGAGAGTAAAAAATGAAAAATGTGCTA 631
DB 537 AAGGACGGTGAAGATCTTAAAGAGCAGCTCAACAGAGAGTAAAAAATGAAAAATGTGCTA 596
QY 632 GGAGAAATCCAGGATGAAGTAGAAAAATCAGAAATCAGTGAAGAACACAGATGCTTTCGGGC 691
DB 597 GGAGAAATCCAGGATGAAGTAGAAAAATCAGAAATCAGTGAAGAACACAGATGCTTTCGGGC 656
QY 692 AAAATAGAGAAATATAATGTTCCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGTGA 751
DB 657 AAAATAGAGAAATATAATGTTCCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGTGA 716
QY 752 CCAACTCAAACTAAGATTTCTCGGGCCCCAAGCGCAAGTGAAGTGAAGAGAGATCTCT 811
DB 717 CCAACTCAAACTAAGATTTCTCGGGCCCCAAGCGCAAGTGAAGTGAAGAGAGATCTCT 776
QY 812 GAAAAACAGCTATTTCTTAGATGACCTGGAATATAGCCAGGTGAGTGTGATCTTTCTACA 871
DB 777 GAAAAACAGCTATTTCTTAGATGACCTGGAATATAGCCAGGTGAGTGTGATCTTTCTACA 836
QY 872 TTGACTCGGAAAAATAGAGATGAGCAATCTGGAATCTTCAACGCTCTTCAAGAAC 931
DB 837 TTGACTCGGAAAAATAGAGATGAGCAATCTGGAATCTTCAACGCTCTTCAAGAAC 896
QY 932 TCTATAAGAGATCGAATGGCCAAAGTACAGGAGCTGTGTCCAAACAAAGCAGCTCAAC 991
DB 897 TCTATAAGAGATCGAATGGCCAAAGTACAGGAGCTGTGTCCAAACAAAGCAGCTCAAC 956
QY 992 AACTATACAAATGAGCTGAAAGCCAGTGTGCGGAAATCNAATTCATAAATCGAGCAA 1051
DB 957 AACTATACAAATGAGCTGAAAGCCAGTGTGCGGAAATCNAATTCATAAATCGAGCAA 1016
QY 1052 AAGGAGATGTGCCCCAGGTCTTGAGGTCTGATCATCCCATCAGGAAGGGGAAAAAGATT 1111
DB 1017 AAGGAGATGTGCCCCAGGTCTTGAGGTCTGATCATCCCATCAGGAAGGGGAAAAAGATT 1076
QY 1112 TCTGCAATAGAGATAGCTCGGAGTCTGTTCCACCCCTCGCGAAGATGATCTCCCGTAC 1171
DB 1077 TCTGCAATAGAGATAGCTCGGAGTCTGTTCCACCCCTCGCGAAGATGATCTCCCGTAC 1136
QY 1172 TCCAGGTTAAGAGTGAAGTTCACAGCTGTGATCCAGGCCACTAAGTCCAGATTC 1231
DB 1137 TCCAGGTTAAGAGTGAAGTTCACAGCTGTGATCCAGGCCACTAAGTCCAGATTC 1196
QY 1232 AGAGCTCCAGTCTTTCTGAAAAATTTCTCTCCCAAGCAATGAAAGAGTTTTCAGGCACCT 1291
DB 1197 AGAGCTCCAGTCTTTCTGAAAAATTTCTCTCCCAAGCAATGAAAGAGTTTTCAGGCACCT 1256
QY 1292 GCAAGAGAGAGCTCGTGGAAATGTGAGAGAGTCTATCCAAATGGAGCGCTCTCTTGGCC 1351
DB 1257 GCAAGAGAGAGCTCGTGGAAATGTGAGAGAGTCTATCCAAATGGAGCGCTCTCTTGGCC 1316
QY 1352 AACCAGAGAGTGTTCACATCAGCTGCTTCCGTTGCTCTTATTCGAAACAAACATCAGT 1411

QY 3572 TTGCTAGTCAATTTATAGAGATTTAGCTTTAATATTTTATAGAGATGTAACAATCTGCT 3631
Db 3537 TTGCTAGTCAATTTATAGAGATTTAGCTTTAATATTTTATAGAGATGTAACAATCTGCT 3596
QY 3632 TTCTTAGTCTTACTAGTCTGAACAATTTTATTCATTAAGAGATTTTAAATTAATTTG 3590
Db 3597 TTCTTAGTCTTACTAGTCTGAACAATTTTATTCATTAAGAGATTTTAAATTAATTTG 3655

RESULT 4
AK000335 3277 bp mRNA linear PRI 22-FEB-2000
LOCUS Homo sapiens cDNA FLJ20328 fis, clone HEP10039.
DEFINITION AK000335
VERSION AK000335.1 GI:7020350
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (sites)
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3277)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission

TITLE Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
JOURNAL University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1,
AUTHORS Minato-Ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- and 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES
source
1. .3277
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEP10039"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_lib="HEP"
/notes="cloning vector pME18SFL3"
179..1978
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAA91092.1"
/db_xref="GI:7020351"
/translations="MNCIGESRHEVEKSEISENTDASQIKYVPLNRLKMFEG
EPQTILRAQSRASGRKISENSYSLDLLEIGPGLSSSTFDSKESNRRLLEPRL
SETIKDRWAKYQAAVKSSSTNYNLEKASGEIKIHMEQKNVPEVCITHQ
EGEIKISANENSLAVRTPAEDSDRSQVSKVQPPVHPKPLSPDSRASSLSSESPKA
MKFQAPARETCVCKTVYPMERLLANQVPHISCFRCSCYNNKLSICTYASLHRI
YCKPHNOLFKSKGNYDEGHRPHKDLWASKNEEILERPAOLANARETPHSQVE
DAPKAVGLAASMEAKASQOEKDKPAETKRLIAMPPTPELSSGSALBEGIKMS
KPKWPDEISKEPEFEDVLDLKURRSSTUKERSRPETVAASQTSVSKPXTVSP
PRKMSMEQSEESVGGVAVRKKQVENAKSKNGNKGKTTWONKESKGETGRSKE
GSLNENENLNGADSDENDFLKQSPQEPKPSLWSSFVDNTFAEEFTTQOKS
QVLEWGEVVEVLEVEQIKNRYDEDEDESE"
BASE COUNT 1089 a 651 c 703 g 834 t
ORIGIN

Query Match 84.0%; Score 3114; DB 9; Length 3277;
Best Local Similarity 99.9%; Pred. No. 0;

Matches 3264; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 439 AGACCATCTCTGCTGAAGTGACAAGCCACGCTGCTTCTTGAGCCAAAGCTGACCAAGA 498
Db 1 AGACCATCTCTGCTGAAGTGACAAGCCACGCTGCTTCTTGAGCCAAAGCTGACCAAGA 60
QY 439 AGAACAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAAGCCCTCTGTTCAAGGTG 558
Db 61 AGAACAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAAGCCCTCTGTTCAAGGTG 120
QY 559 ATATCCCCACATCAAGGAGCGTGAGGATCTTAAAGACCACTCAACAGAAAGTAAATAAT 618
Db 121 ATATCCCCACATCAAGGAGCGTGAGGATCTTAAAGACCACTCAACAGAAAGTAAATAAT 180
QY 619 GGAAAAATTGCTAGGAGAAATCCAGGCATGAAGTAGAAAAATCAGAAATCAGTGAAAAAC 678
Db 181 GGAAAAATTGCTAGGAGAAATCCAGGCATGAAGTAGAAAAATCAGTAATCAGTGAAAAAC 240
QY 679 AGATGCTTCGGGCAAAATAGAGAAATATAATATGTTCCGCTGAACAGGCTTAAGATGATGT 738
Db 241 AGATGCTTCGGGCAAAATAGAGAAATATAATATGTTCCGCTGAACAGGCTTAAGATGATGT 300
QY 739 TGAGAAAGGTGAACCAACTCAACTAAGATTCTCGGGCCCAAGCCGAAAGTGAAGTGG 798
Db 301 TGAGAAAGGTGAACCAACTCAACTAAGATTCTCGGGCCCAAGCCGAAAGTGAAGTGG 360
QY 799 AAGGAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAATAGCCAGGTCACTT 858
Db 361 AAGGAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAATAGCCAGGTCACTT 420
QY 859 GTCATCTTTTACATTTGATCTCGAGAAAAATAGAGTAGAGAAAAATCTGGAATCTTCCACG 918
Db 421 GTCATCTTTTACATTTGATCTCGAGAAAAATAGAGTAGAGAAAAATCTGGAATCTTCCACG 480
QY 919 CCTCTCAGAAACCTCTATAAAGGATCGAATGCGCAAGTACAGGAGCTGTGTCTCAACA 978
Db 481 CCTCTCAGAAACCTCTATAAAGGATCGAATGCGCAAGTACAGGAGCTGTGTCTCAACA 540
QY 979 AAGCAGCTCAACCAACTATACAAATAGCTCAAGAGCCAGTGGTGGGAAATCAAAATCA 1038
Db 541 AAGCAGCTCAACCAACTATACAAATAGCTCAAGAGCCAGTGGTGGGAAATCAAAATCA 600
QY 1039 TAAATGGAGCAAAAGGAGAAATGTGCCCCCAGGTCTCTGAGGTCTGCATCACCCTCAGGA 1098
Db 601 TAAATGGAGCAAAAGGAGAAATGTGCCCCCAGGTCTCTGAGGTCTGCATCACCCTCAGGA 660
QY 1099 AGGGAAAAAGATTTCTGCAAAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCTGCCGAAGA 1158
Db 661 AGGGAAAAAGATTTCTGCAAAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCTGCCGAAGA 720
QY 1159 TGACTCCCGTCACTCCAGGTTAAGAGTAGGTTCAACAGCCTGTCCATCCCAAGCCACT 1218
Db 721 TGACTCCCGTCACTCCAGGTTAAGAGTAGGTTCAACAGCCTGTCCATCCCAAGCCACT 780
QY 1219 AAGTCCAGATTCAGAGGCTCCAGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAA 1278
Db 781 AAGTCCAGATTCAGAGGCTCCAGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAA 840
QY 1279 GTTTCAGCACCTCTCAAGAGAGACCTCGTGGAAATGTGAGAAAGACAGTCTATCCAATGGA 1338
Db 841 GTTTCAGCACCTCTCAAGAGAGACCTCGTGGAAATGTGAGAAAGACAGTCTATCCAATGGA 900
QY 1339 CGCTCTCTTGGCCAAACAGCAGGTTTTCACATCAGTGTCTCCGTTGCTCTCTATTGCAA 1398
Db 901 CGCTCTCTTGGCCAAACAGCAGGTTTTCACATCAGTGTCTCCGTTGCTCTCTATTGCAA 960
QY 1399 CAAACAACTCAGTCTAGGACATATGATCTTTACATGGAAGATCTTATGTAAGCTCA 1458
Db 961 CAAACAACTCAGTCTAGGACATATGATCTTTACATGGAAGATCTTATGTAAGCTCA 1020
QY 1459 CTTCAATCAACTCTTTTAAATCTAAGGCAACTATGATGAAGGCTTTTGGGCACAGCCACA 1518
Db 1021 CTTCAATCAACTCTTTTAAATCTAAGGCAACTATGATGAAGGCTTTTGGGCACAGCCACA 1080

QY 1519 CAAGGATCTATGGGCAAGCAGAAATGAAACGAGAGATTTTGGAGAGACCAAGCCAGCT 1578
Db 1081 CAAGGATCTATGGGCAAGCAGAAATGAAACGAGAGATTTTGGAGAGACCAAGCCAGCT 1140
QY 1579 TGCAAATGCAAGGGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTATTGCTAAAGGT 1638
Db 1141 TGCAAATGCAAGGGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTATTGCTAAAGGT 1200
QY 1639 GGGTGTCTGGTGTGCAAGTATGGAAGCCAGGCCCTCTCTCAGCAGGAGAAAGAAAGCA 1698
Db 1201 GGGTGTCTGGTGTGCAAGTATGGAAGCCAGGCCCTCTCTCAGCAGGAGAAAGAAAGCA 1260
QY 1699 GCCAGCTGAAACCAAGAGCTGAGGATGCCCTGGCCACCCCTCAGTAACTTGGAGATTC 1758
Db 1261 GCCAGCTGAAACCAAGAGCTGAGGATGCCCTGGCCACCCCTCAGTAACTTGGAGATTC 1320
QY 1759 AGGAAGTCTGGAGGAGAGGATCAAAATGTCAAAGCCCAATGSCCTCTCTGAAAGAGA 1818
Db 1321 AGGAAGTCTGGAGGAGAGGATCAAAATGTCAAAGCCCAATGSCCTCTCTGAAAGAGA 1380
QY 1819 AATCAGCAAGCCGGAAGTCTCTGAGGATGTGATCTGATCTGAAAGAGCTTAAGACGATC 1878
Db 1381 AATCAGCAAGCCGGAAGTCTCTGAGGATGTGATCTGATCTGAAAGAGCTTAAGACGATC 1440
QY 1879 TTCTTCACTGAAAGGAAAGAGCCGCCATTTCACTGTAGCAGCTTCATTTCAAAGCACCTC 1938
Db 1441 TTCTTCACTGAAAGGAAAGAGCCGCCATTTCACTGTAGCAGCTTCATTTCAAAGCACCTC 1500
QY 1939 TGTCAAGACCCCAAAACTGTGTCCCCACTATCAGGAAGGCTGGAGCATGTCAAGCA 1998
Db 1501 TGTCAAGACCCCAAAACTGTGTCCCCACTATCAGGAAGGCTGGAGCATGTCAAGCA 1560
QY 1999 GAGTGAAGAGTCTGTGGTGGAGAGTGTGCAAGAGGAAACAAAGTGGAAATGCCCAAGGC 2058
Db 1561 GAGTGAAGAGTCTGTGGTGGAGAGTGTGCAAGAGGAAACAAAGTGGAAATGCCCAAGGC 1620
QY 2059 TTCTAAGAGAGATGGGAATGTGGGAAAAACACCTGGCAAAAACAAAGATCTAAAGGAGA 2118
Db 1621 TTCTAAGAGAGATGGGAATGTGGGAAAAACACCTGGCAAAAACAAAGATCTAAAGGAGA 1680
QY 2119 GACAGGAGAGAGATGAAGAGGTCATAGTTTGGAGATGGAGATGGAATCTTTGTAGA 2178
Db 1681 GACAGGAGAGAGATGAAGAGGTCATAGTTTGGAGATGGAGATGGAATCTTTGTAGA 1740
QY 2179 AAATGGTCAGACTCCGATGAAGATGATAACAGCTTCTCAACACACATCTCCACAGA 2238
Db 1741 AAATGGTCAGACTCCGATGAAGATGATAACAGCTTCTCAACACACATCTCCACAGA 1800
QY 2239 ACCCAAGTCTCTGAATTTGTCGAGTTTGTAGACAAACACCTTTGCTGAAGAAATTCACCTAC 2298
Db 1801 ACCCAAGTCTCTGAATTTGTCGAGTTTGTAGACAAACACCTTTGCTGAAGAAATTCACCTAC 1860
QY 2299 TCAGAAATCAGAAATCCAGGATGTGAACTCTGGGAGGGAGAGTGGTCAAAGAGCTCTC 2358
Db 1861 TCAGAAATCAGAAATCCAGGATGTGAACTCTGGGAGGGAGAGTGGTCAAAGAGCTCTC 1920
QY 2359 TGTGGAAGACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGACA 2418
Db 1921 TGTGGAAGACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGACA 1980
QY 2419 AAATGCAATGATGCTGGGCTTAAATTCATGTTAGTGTAGGAGCCAGCTGCCCTTTGTC 2478
Db 1981 AAATGCAATGATGCTGGGCTTAAATTCATGTTAGTGTAGGAGCCAGCTGCCCTTTGTC 2040
QY 2479 AAAATGTGATGACATGAAGCAGGATCCAGCATGAAATGTAATTTACTTTGGAAGTAACT 2538
Db 2041 AAAATGTGATGACATGAAGCAGGATCCAGCATGAAATGTAATTTACTTTGGAAGTAACT 2100
QY 2539 TTGGAAGAGAAATCTCTTTAAATCAAAAACAAACCAAAAAACAAAAACACATTC 2598
Db 2101 TTGGAAGAGAAATCTCTTTAAATCAAAAACAAACCAAAAAACAAAAACACATTC 2160

QY 2599 TAAATCTAGAGATAAATTTACTTAAATTTCTTCAATTTTAGCAGTGATGATATGCAATAGT 2658
Db 2161 TAAATCTAGAGATAAATTTACTTAAATTTCTTCAATTTTAGCAGTGATGATATGCAATAGT 2220
QY 2659 GCTGTAAAGGCTTTAACTGGGGAATATTCACCTGTATATAGCCAGATTTCTACTGTAT 2718
Db 2221 GCTGTAAAGGCTTTAACTGGGGAATATTCACCTGTATATAGCCAGATTTCTACTGTAT 2280
QY 2719 TCCAAAAGGCAATATTAAAGGTAGATAGATGATATTAGTATATTTTGTACACACTATTTT 2778
Db 2281 TCCAAAAGGCAATATTAAAGGTAGACAGATGATATTAGTATATTTTGTACACACTATTTT 2340
QY 2779 GGAATTTAGAGAACATACAGAAAGGAATTTAGGGGCTTAAACATTACGACTGAAATGCATTT 2838
Db 2341 GGAATTTAGAGAACATACAGAAAGGAATTTAGGGGCTTAAACATTACGACTGAAATGCATTT 2400
QY 2839 AGTATAAGGSCACAGTTGTATATTTTAAATCAATACCAATTTTAAATTTTAAATTTT 2898
Db 2401 AGTATAAGGSCACAGTTGTATATTTTAAATCAATACCAATTTTAAATTTTAAATTTT 2460
QY 2899 ACCTGTTAAAGAGATTTTACTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2958
Db 2461 ACCTGTTAAAGAGATTTTACTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2520
QY 2959 TATCAGGAATTTACTACTTTATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3018
Db 2521 TATCAGGAATTTACTACTTTATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2580
QY 3019 GAGGTATAATCAACAGAGACACTTTTGGAGCAATTTGAAACCAACCTACACTCTTCGG 3078
Db 2581 GAGGTATAATCAACAGAGACACTTTTGGAGCAATTTGAAACCAACCTACACTCTTCGG 2640
QY 3079 TGTCTAGAGAGATCTGTCTCTCCCAATTAAGCTTTTGTATCTGCCAGTGAATTTACTGT 3138
Db 2641 TGTCTAGAGAGATCTGTCTCTCCCAATTAAGCTTTTGTATCTGCCAGTGAATTTACTGT 2700
QY 3139 ACTCCAAATGATTTCTTTCTGCTGATATCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 3198
Db 2701 ACTCCAAATGATTTCTTTCTGCTGATATCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 2760
QY 3199 GCAATATTTTAGTAAATACCTTCGGGATCACTGTCCTCCCTCTCTCTCTCTCTCTCTCT 3258
Db 2761 GCAATATTTTAGTAAATACCTTCGGGATCACTGTCCTCCCTCTCTCTCTCTCTCTCTCT 2820
QY 3259 GAAGAGTTTAAAGGAGAGAGAGAAAGAACTGTCTTACACCACTTGAGCTCAGACCTCTA 3318
Db 2821 GAAGAGTTTAAAGGAGAGAGAGAGAAAGAACTGTCTTACACCACTTGAGCTCAGACCTCTA 2880
QY 3319 AACCTGTATTTCTCTTATGATGTCCTCTTTTGGAGACACTTAAATTTTAAATTTTAAAT 3378
Db 2881 AACCTGTATTTCTCTTATGATGTCCTCTTTTGGAGACACTTAAATTTTAAATTTTAAAT 2940
QY 3379 GCTCTGAAATATATTGATTTTATCAGATTTCTCAGGGTGAATTTAAACCAACTATAG 3438
Db 2941 GCTCTGAAATATATTGATTTTATCAGATTTCTCAGGGTGAATTTAAACCAACTATAG 3000
QY 3439 GCCTTTTCTTGGGATGATTTCTTCTAGTCTTAAAGTTTGGGACATTTAAACCTTGAGTAC 3498
Db 3001 GCCTTTTCTTGGGATGATTTCTTCTAGTCTTAAAGTTTGGGACATTTAAACCTTGAGTAC 3060
QY 3499 ATTTGTTGTACACAGTTGATTTCCAAATTTGTATGATGGAGGAGGAGGTGCTTAAAGC 3558
Db 3061 ATTTGTTGTACACAGTTGATTTCCAAATTTGTATGATGGAGGAGGAGGTGCTTAAAGC 3120
QY 3559 TGTAAGGCTTTCTTTGACTGCAATTTATAGAGATTTAGCTTTTAAATTTTAAATTTTAAATTT 3618
Db 3121 TGTAAGGCTTTCTTTGACTGCAATTTATAGAGATTTAGCTTTTAAATTTTAAATTTTAAATTT 3180
QY 3619 AAAACATCTCTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3678
Db 3181 AAAACATCTCTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3240
QY 3679 AATTAATTTTGAAGAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 3705

Db 3241 AATTAAATTTGAAAAAATAAAAAA 3267

RESULT 5

AF198455

LOCUS AF198455 3550 bp mRNA linear PRI 10-JAN-2000

DEFINITION Homo sapiens epithelial protein lost in neoplasm alpha (EPLIN)

ACCESSION AF198455

VERSION AF198455

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3550)

AUTHORS Maul, R. S. and Chang, D. D.

TITLE EPLIN, epithelial protein lost in neoplasm

JOURNAL Oncogene 18 (54), 7838-7841 (1999)

MEDLINE 20087188

PUBMED 10618726

REFERENCE 2 (bases 1 to 3550)

AUTHORS Maul, R. S. and Chang, D. D.

TITLE Direct Submission

JOURNAL Submitted (25-OCT-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA

FEATURES

source 1..3550

gene 1..3550

CDS 474..2276

1 /organism="Homo sapiens"

2 /mol_type="mRNA"

3 /db_xref="taxon:9606"

4 /chromosome="12"

5 /map="12q13"

6 /protein_id="AA233756.1"

7 /codon_start=1

8 /product="epithelial protein lost in neoplasm alpha"

9 /protein_id="AA233756.1"

10 /db_xref="GI:6685008"

11 /translations="MENCIGESRHEVEKSEISNTASGKIEKYNVPLNRLKMFEG
EPQTILRAQSSAGSRKISNSYSLDLEIGFGLSSSTFSEKNERRNLELPRL
SETSIDRMAKQAAVSQSSSTYNELKASGGEIKIHKEOKENVVPPQVPCITHQ
EGEKISANENSLAVRSTPAEDDSDPSGVQVQHPKPLSPDSRSLSESPPK
AMKFOAPARETCVEQKTYYPMERLLANQVPHISCFRCSVCNKLKSLGTYSUHRG
IYCKPFPNQLFKAGNIDEGFGRPHDKLWASKNEEILERPQLANARETPHSEGV
EDAPIAKGVLAASMEAKSSQKEDKPAETKKLRIAWPPPTLGSQSALEEGIM
SKPWPEDDISKPEVEDYLDLKLRRSSSLKRSRPTVAASQSTSVKSPKTVS
PPIRKGMSEQSEESVGGVABRKOVENAKSKKNVGTWQNKESKGETGRSK
EGHLEMNENLVENGADSDSDNSFLKQSPQEPKSLNWSFVDNTPAEETQNKQK
SQDELWEGEVVKELSVEEQIKENRYDEDEEE"

polyA_signal 3527..3532

BASE COUNT 1153 a 716 c 773 g 908 t

ORIGIN

Query Match 83.68; Score 3098; DB 9; Length 3550;

Best Local Similarity 99.94; Pred. No. 0;

Matches 3388; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 302 AACACCCAAAATCTCTCCAGCAGCTTTAGAAAAGGGACCTGACTGTGTTAAAGAGAAG 361

DB 159 AACACCCAAAATCTCTCCAGCAGCTTTAGAAAAGGGACCTGACTGTGTTAAAGAGAAG 218

QY 362 TGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGCTCTCTACGGAAACAGCAGCACT 421

DB 219 TGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGCTCTCTACGGAAACAGCAGCACT 278

QY 422 GAGATTAGGCACAGACAGACCATCTCTCTGCTGAAGTGACAGCACCGCTGCTCTGGA 481

DB 279 GAGATTAGGCACAGACAGACCATCTCTCTGCTGCTGAAGTGACAGCACCGCTGCTCTGGA 338

QY 482 GCCAAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGACTCAGGTCTCCTCTGAA 541

DB 339 GCCAAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGACTCAGGTCTCCTCTGAA 398

QY 542 GGCCTCGTTGAGGTCGATATCCCCACATCAAGGACGCTGAGGATCTTAAAGACCACTCA 601

DB 399 GGCCTCGTTGAGGTCGATATCCCCACATCAAGGACGCTGAGGATCTTAAAGACCACTCA 458

QY 602 ACAGAAAGTAAAAAAATGAAAAATTTGTCTAGGAGAAATCCAGGCATGAAAGTAAAGAAATCA 661

DB 459 ACAGAAAGTAAAAAAATGAAAAATTTGTCTAGGAGAAATCCAGGCATGAAAGTAAAGAAATCA 518

QY 662 GAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAAC 721

DB 519 GAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAAC 578

QY 722 AGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACTAAGATTTCTCGGCGCCAA 781

DB 579 AGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACTAAGATTTCTCGGCGCCAA 638

QY 782 AGCCGAAGTGAAGTGAAGGAAGATCTCTGAAAACAGCTATTCTCTAGATGACCTGAA 841

DB 639 AGCCGAAGTGAAGTGAAGGAAGATCTCTGAAAACAGCTATTCTCTAGATGACCTGAA 698

QY 842 ATAGGCCAGGTCAGTTGCTGCTTCTATCTTGAATTCGAGAGAAAAATGAGAGTAGACGA 901

DB 699 ATAGGCCAGGTCAGTTGCTGCTTCTATCTTGAATTCGAGAGAAAAATGAGAGTAGACGA 758

QY 902 AATCTGGAACCTCCAGGCTCTCAGAAAACCTCTATAAAGGATCGAATGGCCAAAGTACCAG 961

DB 759 AATCTGGAACCTCCAGGCTCTCAGAAAACCTCTATAAAGGATCGAATGGCCAAAGTACCAG 818

QY 962 GCAGCTGTGTCGAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT 1021

DB 819 GCAGCTGTGTCGAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT 878

QY 1022 GCGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCCTCAGGTC 1081

DB 879 GCGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCCTCAGGTC 938

QY 1082 TGCATCACCCATCAGAAAGGGAAAAAGATTTCTGCAAAATGAGAAATAGCTGGCAGTCCGT 1141

DB 939 TGCATCACCCATCAGAAAGGGAAAAAGATTTCTGCAAAATGAGAAATAGCTGGCAGTCCGT 998

QY 1142 TCCACCCCTGCGGAAGATGACTCC --GTGACTCCAGGTTAAGAGTGAAGTTCACAG 1198

DB 999 TCCACCCCTGCGGAAGATGACTCCCGAGTGAAGTTCAGAGTGAAGTTCACAG 1058

QY 1199 CCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCCAGTCTTTCTGAAAGTTCT 1258

DB 1059 CCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCCAGTCTTTCTGAAAGTTCT 1118

QY 1259 CCTCCCAAGCAATGAAGAAAGTTTTCAGGCACCTCGAAGAGAGCTGCGTGAATGTCTGAG 1318

DB 1119 CCTCCCAAGCAATGAAGAAAGTTTTCAGGCACCTCGAAGAGAGCTGCGTGAATGTCTGAG 1178

QY 1319 AAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCAAACAGCAGGTTGTTTACATCAGTGC 1378

DB 1179 AAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCAAACAGCAGGTTGTTTACATCAGTGC 1238

QY 1379 TTCGTTGCTCTTATGTCGAAACCAAACTCAGTCTAGGAAACATATGCATCTTTACATGGA 1438

DB 1239 TTCGTTGCTCTTATGTCGAAACCAAACTCAGTCTAGGAAACATATGCATCTTTACATGGA 1298

QY 1439 AGAATCTATTGTAAGCTCAGTCTCAATCAACTCTTTTAAATCTAAGGGCACTATGATGAA 1498

DB 1299 AGAATCTATTGTAAGCTCAGTCTCAATCAACTCTTTTAAATCTAAGGGCACTATGATGAA 1358

QY 1499 GCGTTTGGGCACAGACCAACAGGATCTATGGCAAGCAAAATGAAACAGCAAGAGATT 1558

DB 1359 GCGTTTGGGCACAGACCAACAGGATCTATGGCAAGCAAAATGAAACAGCAAGAGATT 1418

AX017487
LOCUS AX017487 2783 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 31 from Patent WO9947655.
ACCESSION AX017487
VERSION AX017487.1 GI:10042284
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Schmitt, A., Specht, T., Dahl, E., Hinzmann, B., Rosenthal, A. and
Pilarczyk, C.
TITLE Human nucleic acid sequences from normal breast tissue
JOURNAL Patent: WO 9947655-A 31 23-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..2783
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 901 a 526 c 609 g 747 t
ORIGIN
Query Match 71.4%; Score 2647; DB 6; Length 2783;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2697; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1001 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTTATAAATGGAGCAAAAGAGAAAT 1060
DB 40 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTTATAAATGGAGCAAAAGAGAAAT 99
QY 1061 GTGCCCCCAGGTCCTGAGTCTGCATACCCATCAGGAGGGGAAAGATTTCTGCAAAAT 1120
DB 100 GTGCCCCCAGGTCCTGAGTCTGCATACCCATCAGGAGGGGAAAGATTTCTGCAAAAT 159
QY 1121 GAGAATAGCCTGGCAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTGACTCCACAGTT 1180
DB 160 GAGAATAGCCTGGCAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTGACTCCACAGTT 219
QY 1181 AAGAGTGAAGTTCAACAGCTTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCCTCC 1240
DB 220 AAGAGTGAAGTTCAACAGCTTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCCTCC 279
QY 1241 AGTCCTTCTGAAGTCTCTCCCAAGCAATGAAGATTTTCAGGACCTTGCAGAGAG 1300
DB 280 AGTCCTTCTGAAGTCTCTCCCAAGCAATGAAGATTTTCAGGACCTTGCAGAGAG 339
QY 1301 ACCTCGTGGAAATGTCAAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCCAACCCAGAG 1360
DB 340 ACCTCGTGGAAATGTCAAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCCAACCCAGAG 399
QY 1361 GTGTTTTCATCAGTGTGTCGTTCCATTCCTATTCGCAACAACTCAGTCTAGGAACA 1420
DB 400 GTGTTTTCATCAGTGTGTCGTTCCATTCCTATTCGCAACAACTCAGTCTAGGAACA 459
QY 1421 TATGCATCTTTACATGGAGATCTATTGTAGGCTCATTCAATCAACTCTTTAAATCT 1480
DB 460 TATGCATCTTTACATGGAGATCTATTGTAGGCTCATTCAATCAACTCTTTAAATCT 519
QY 1481 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCTATGGCGAAGCAAA 1540
DB 520 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCTATGGCGAAGCAAA 579
QY 1541 AATGAAACGAGAGATTTGGAGAGACCAGCCAGCTTGCAAATGCAAGGGAGACCCCT 1600
DB 580 AATGAAACGAGAGATTTGGAGAGACCAGCCAGCTTGCAAATGCAAGGGAGACCCCT 639
QY 1601 CACAGCCAGGGGTAGAGATGCCCTTATTCGTAAGGTGGGTGTCCTGGCTGCAAGTATG 1660
DB 640 CACAGCCAGGGGTAGAGATGCCCTTATTCGTAAGGTGGGTGTCCTGGCTGCAAGTATG 699

QY 1661 GAAGCCCAAGCCCTCTCTCTCAGCAGGAGAAAGCAAGCCAGCTGAAACCAAGAGCTG 1720
DB 700 GAAGCCCAAGCCCTCTCTCTCAGCAGGAGAAAGCAAGCCAGCTGAAACCAAGAGCTG 759
QY 1721 AGATTCGCTTGGCCACCCCTCCTGAAAGTTTCAGGAAGTGCCTTTGGAGGAAGGG 1780
DB 760 AGATTCGCTTGGCCACCCCTCCTGAAAGTTTCAGGAAGTGCCTTTGGAGGAAGGG 819
QY 1781 ATCAAAATGTCAAAAGCCCAATGSCCTCCTGAAGAGCAAAATCAGCAAGCCCGAAGTTCT 1840
DB 820 ATCAAAATGTCAAAAGCCCAATGSCCTCCTGAAGAGCAAAATCAGCAAGCCCGAAGTTCT 879
QY 1841 GAGGATCTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAAGAGC 1900
DB 880 GAGGATCTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAAGAGC 939
QY 1901 CGCCCATCTACTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAATCTGTG 1960
DB 940 CGCCCATCTACTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAATCTGTG 999
QY 1961 TCCCCACCTATCAGGAAGGCTCGACATGTCTAGACAGAGTCAAGAGTCTGTGGGTGGA 2020
DB 1000 TCCCCACCTATCAGGAAGGCTCGACATGTCTAGACAGAGTCAAGAGTCTGTGGGTGGA 1059
QY 2021 AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGCTTCTTAAGAGAAATGGAAATGTG 2080
DB 1060 AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGCTTCTTAAGAGAAATGGAAATGTG 1119
QY 2081 GGAAAAACAACTGGGCAAAACAAAGAAATCTTAAAGGAGACAGGGAAGAGAACTAAGGAA 2140
DB 1120 GGAAAAACAACTGGGCAAAACAAAGAAATCTTAAAGGAGACAGGGAAGAGAACTAAGGAA 1179
QY 2141 GGTCTATGTTGGAGATGGAGAAATGAGAAATCTTGTAGAAATGGTGCAGACTCCGATGAA 2200
DB 1180 GGTCTATGTTGGAGATGGAGAAATGAGAAATCTTGTAGAAATGGTGCAGACTCCGATGAA 1239
QY 2201 GATGATAACAGCTTCTCBAACCAACATCTCCACAGAACCCCAAGTCTCTGAATGGTGC 2260
DB 1240 GATGATAACAGCTTCTCBAACCAACATCTCCACAGAACCCCAAGTCTCTGAATGGTGC 1299
QY 2261 AGTTTTGTAGACAAACACCTTTGCTGAAGAAATTCACCTACTCAGAAATCAGAAATCCCAGGAT 2320
DB 1300 AGTTTTGTAGACAAACACCTTTGCTGAAGAAATTCACCTACTCAGAAATCAGAAATCCCAGGAT 1359
QY 2321 GTGGAACCTGTGGAGGAGAGTGGTCAAAGAGCTCTCTGTGGAAAGACAGATAAAGAGA 2380
DB 1360 GTGGAACCTGTGGAGGAGAGTGGTCAAAGAGCTCTCTGTGGAAAGACAGATAAAGAGA 1419
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAAGTGAACAAATTCGAATGATGCTGGGCCCTT 2440
DB 1420 AATCGGTATTATGATGAGGATGAGGATGAAGTGAACAAATTCGAATGATGCTGGGCCCTT 1479
QY 2441 AAATTCATGTTAGTGTAGCGGCCACTGCCCTTTGTCAAAATGTGATGCACATAGCAG 2500
DB 1480 AAATTCATGTTAGTGTAGCGGCCACTGCCCTTTGTCAAAATGTGATGCACATAGCAG 1539
QY 2501 GTATCCAGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAAAGAAATTCCTTTCTTAA 2560
DB 1540 GTATCCAGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAAAGAAATTCCTTTCTTAA 1599
QY 2561 AATCAAAACAAACAAAAAAACACAAAAACACAAATCTTAAATCTAGAGATAAATCTTAC 2620
DB 1600 AATCAAAACAAACAAAAAAACACAAAAACACAAATCTTAAATCTAGAGATAAATCTTAC 1659
QY 2621 TTAATTTCTTCTTTTAGCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 2680
DB 1660 TTAATTTCTTCTTTTAGCAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 1719
QY 2681 AAATATTCACCTGTATAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAAGT 2740
DB 1720 AAATATTCACCTGTATAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAAGT 1779

Db 520 AAGGCAACTATGATGAAGCTTTGGGCACAGACCACACAAGGATCTATGGGCAAGCMAA 579
Qy 1541 AATGAAAACGAAGAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGAGAGACCCCT 1600
Db 580 AATGAAAACGAAGAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGAGAGACCCCT 639
Qy 1601 CACAGCCAGGGGTAGAGATGCCCTATTGCTAAGGTGGGTGCTCCTGGCTGCAAGTATG 1660
Db 640 CACAGCCAGGGGTAGAGATGCCCTATTGCTAAGGGGGGTGCTCCTGGCTGCAAGTATG 699
Qy 1661 GAAGCCAAAGCCTCCTCTCAGCAGGAGAGAGAAAGCAAGCCAGCTGAAACCAAGAAAGCTG 1720
Db 700 GAAGCCAAAGCCTCCTCTCAGCAGGAGAGAGAAAGCAAGCCAGCTGAAACCAAGAAAGCTG 759
Qy 1721 AGSATGCCCTGGCCACCCCTCAGTGAACCTTGGAAGTTTCAGAAAGTGCCTTTGGAGGAAGG 1780
Db 760 AGSATGCCCTGGCCACCCCTCAGTGAACCTTGGAAGTTTCAGAAAGTGCCTTTGGAGGAAGG 819
Qy 1781 ATCAAAATGTCAAAAGCCCAATGCCCTCTGAGAGAGAAATCAGCAAGCCCGAAGTTCTCT 1840
Db 820 ATCAAAATGTCAAAAGCCCAATGCCCTCTGAGAGAGAAATCAGCAAGCCCGAAGTTCTCT 879
Qy 1841 GAGGATGCTGATCTAGATCTGAAGAGCTAAGACGATCTTTCTTCACTGAAGGAAGAAAGC 1900
Db 880 GAGGATGCTGATCTAGATCTGAAGAGCTAAGACGATCTTTCTTCACTGAAGGAAGAAAGC 939
Qy 1901 CGCCCATCTAGTGAAGCTTCAATTTCAAGACCTCTCTCAAGAGCCCAAAATCTGTG 1960
Db 940 CGCCCATCTAGTGAAGCTTCAATTTCAAGACCTCTCTCAAGAGCCCAAAATCTGTG 999
Qy 1961 TCCCACTCTATCAGGAAGGCTGGAGCATGTCAGAGAGAGTGAAGAGTCTGTGGGTGA 2020
Db 1000 TCCCACTCTATCAGGAAGGCTGGAGCATGTCAGAGAGAGTGAAGAGTCTGTGGGTGA 1059
Qy 2021 AGAGTTGAGAAAGGAAACAAGTGGAAATTCGCAAGCTTCTAAGAGAAATGGAATGTG 2080
Db 1060 AGAGTTGAGAAAGGAAACAAGTGGAAATTCGCAAGCTTCTAAGAGAAATGGAATGTG 1119
Qy 2081 GGAAAAACAACCTGGCAAAACAAGAAATCTAAAGGAGAGACAGGGAAGAGAAAGGAA 2140
Db 1120 GGAAAAACAACCTGGCAAAACAAGAAATCTAAAGGAGAGACAGGGAAGAGAAAGGAA 1179
Qy 2141 GGTCAATAGTTGGAGATGGAGAAATGAGAAATCTGTGAAAAATGGTGCAGACTCCGATGA 2200
Db 1180 GGTCAATAGTTGGAGATGGAGAAATGAGAAATCTGTGAAAAATGGTGCAGACTCCGATGA 1239
Qy 2201 GATGATAACAGCTTCTCACAACAACTCTCACAAGACCCCAAGTCTCTGAATGGTGC 2260
Db 1240 GATGATAACAGCTTCTCACAACAACTCTCACAAGACCCCAAGTCTCTGAATGGTGC 1299
Qy 2261 AGTTTGTAGACAACACCTTTGCTGAAGAAATCACTACTCAGAAATCAGAAATCCCAAGAT 2320
Db 1300 AGTTTGTAGACAACACCTTTGCTGAAGAAATCACTACTCAGAAATCAGAAATCCCAAGAT 1359
Qy 2321 GTGGAATCTGGAGGAGAGAGTGGTCAAGAGCTCTGTGGAAGAAACAGATAAAGAGA 2380
Db 1360 GTGGAATCTGGAGGAGAGAGTGGTCAAGAGCTCTGTGGAAGAAACAGATAAAGAGA 1419
Qy 2381 AATCGGTATTATGATGAGGATGAGGATGAGAGTGACAAATTCGAATGATGCTGGGCCCTT 2440
Db 1420 AATCGGTATTATGATGAGGATGAGGATGAGAGTGACAAATTCGAATGATGCTGGGCCCTT 1479
Qy 2441 AAATTCATGTTAGTGTAGGAGCCACTGCCCTTTGTCAAAATGTGATGCAATAGAGAG 2500
Db 1480 AAATTCATGTTAGTGTAGGAGCCACTGCCCTTTGTCAAAATGTGATGCAATAGAGAG 1539
Qy 2501 GTATCCAGCATGAATGTAAATTAAGTGAAGTAACTTTGGAAAAAGAAATCTCTCTTAA 2560
Db 1540 GTATCCAGCATGAATGTAAATTAAGTGAAGTAACTTTGGAAAAAGAAATCTCTCTTAA 1599
Qy 2561 AATCAAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAC 2620
Db 1600 AATCAAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAC 1659

Qy 2621 TTAAATCTCTTCAATTTAGCAGTGATGATATCATAAAGTGTGTAAAGCTTGTAACTGGG 2680
Db 1660 TTAAATCTCTTCAATTTAGCAGTGATGATATCATAAAGTGTGTAAAGCTTGTAACTGGG 1719
Qy 2681 AATATATCCACTGATATAGCCAGATTTCTACTGTATTTCCCAAAAGCAATATTAAGGT 2740
Db 1720 AATATATCCACTGATATAGCCAGATTTCTACTGTATTTCCCAAAAGCAATATTAAGGT 1779
Qy 2741 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAAACATACAGAG 2800
Db 1780 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAAACATACAGAG 1839
Qy 2801 GAATTTAGGGCTTAAACATTTACGACTGAATGCACTTTTAGTATAAAGGGCAGTTTGT 2860
Db 1840 GAATTTAGGGCTTAAACATTTACGACTGAATGCACTTTTAGTATAAAGGGCAGTTTGT 1899
Qy 2861 TATTTTAAATGAATACCAATTTAATTTTGTAGTATTTTACCTGTTTAAAGAGATTTTGT 2920
Db 1900 TATTTTAAATGAATACCAATTTAATTTTGTAGTATTTTACCTGTTTAAAGAGATTTTGT 1959
Qy 2921 CTTTAAATTTTGTAGTTAAATTTCTCTGCTGATATATATGAGAAATTTACTACTTTAT 2980
Db 1960 CTTTAAATTTTGTAGTTAAATTTCTCTGCTGATATATATGAGAAATTTACTACTTTAT 2019
Qy 2981 GTCTGCTCTCTAAACTACATCTGAACCTGAACTCGACGCTCTCTGAGGTATAATACACAGAGAC 3040
Db 2020 GTCTGCTCTCTAAACTACATCTGAACCTCGACGCTCTCTGAGGTATAATACACAGAGAC 2079
Qy 3041 TTTTGTAGGCAATGAAAAACCAACCTACCTCTCTGGTGTGTAGAGAGATCTGCTGTCT 3100
Db 2080 TTTTGTAGGCAATGAAAAACCAACCTACCTCTCTGGTGTGTAGAGAGATCTGCTGTCT 2139
Qy 3101 CCCAAATGAAGCTTTGTATCTGCCAGTGAATTTACTGTACTTCCAAATGATGCTTCTTT 3160
Db 2140 CCCAAATGAAGCTTTGTATCTGCCAGTGAATTTACTGTACTTCCAAATGATGCTTCTTT 2199
Qy 3161 TCTGTGTATCTGTGCTTCTCATAATTAAGTGAAGTGTCAATATTTTAGTAATACCTTC 3220
Db 2200 TCTGTGTATCTGTGCTTCTCATAATTAAGTGTCAATATTTTAGTAATACCTTC 2259
Qy 3221 GGGATCACTGTCCCCCATCTTTCCGTGTTAGACAAAGTGAAGAGTTTAAAGGAGAGAA 3280
Db 2260 GGGATCACTGTCCCCCATCTTTCCGTGTTAGACAAAGTGAAGAGTTTAAAGGAGAGAA 2319
Qy 3281 GAAAGAACTGTCTTACACCACTGTGAGCTCAGACCTTAAACCCCTGTATTTCCCTTATGAT 3340
Db 2320 GAAAGAACTGTCTTACACCACTGTGAGCTCAGACCTTAAACCCCTGTATTTCCCTTATGAT 2379
Qy 3341 GTCCCTTTTGTAGACACTAATTTTAAATACTTACTAGCTCTGAAATATATGATTTTT 3400
Db 2380 GTCCCTTTTGTAGACACTAATTTTAAATACTTACTAGCTCTGAAATATATGATTTTT 2439
Qy 3401 ATCAGATATTTCTCAGGCTGAAATTAACCAACTATAGGCCCTTTTCTGGGATGATTTT 3460
Db 2440 ATCAGATATTTCTCAGGCTGAAATTAACCAACTATAGGCCCTTTTCTGGGATGATTTT 2499
Qy 3461 CTAGCTTTAAGGTTGGGACATTTATAAATGTAGTACATTTGTTGTACACAGTTGATAT 3520
Db 2500 CTAGCTTTAAGGTTGGGACATTTATAAATGTAGTACATTTGTTGTACACAGTTGATAT 2559
Qy 3521 TCCAAATGTATGATGGGAGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTGTACTGC 3580
Db 2560 TCCAAATGTATGATGGGAGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTGTACTGC 2619
Qy 3581 ATTTATAGAGATTTAGCTTTTAAATTTTTTAGAGATGTAAACCAATCTGCTTCTTAGTC 3640
Db 2620 ATTTATAGAGATTTAGCTTTTAAATTTTTTAGAGATGTAAACCAATCTGCTTCTTAGTC 2679
Qy 3641 TTACTAGTCTGAACAACTTTTATTTCAATAAGATTTTAAATTTTAAATTTTGAAGAAAA 3698
Db 2680 TTACTAGTCTGAACAACTTTTATTTCAATAAGATTTTAAATTTTAAATTTTGAAGAAAA 2737

Qy	123	TGTCGTAGACAAGATGGAAATCATCTCCATTTAATAGACGGCAATAGGACCTCACTATCAT	181
Db	71	TGTCGTAGACAAGATGGAAATCATCTCCATTTAATAGACGGCAATAGGACCTCACTATCAT	130
Qy	183	TGAGGGTAAACAGCCAAAGAACTTTCTCTTGTCACAGAAACAAGTCATCGGCTATTGTGG	242
Db	131	TGAGGGTAAACAGCCAAAGAACTTTCTCTTGTCACAGAAACAAGTCATCGGCTATTGTGG	190
Qy	243	AAATATTCTCCAAGTACCAGAAACGACTGAAGAAACAAACATCGAGAGAAGAGAAAGTA	302
Db	191	AAATATTCTCCAAGTACCAGAAACGACTGAAGAAACAAACATCGAGAGAAGAGAAAGTA	250
Qy	303	ACACCGAAAAATCTCTCCAGCACCTTTAGAAAGGGGACCCCTCACTGTGTGTTAAAGAAAGAAGT	362
Db	251	ACACCGAAAAATCTCTCCAGCACCTTTAGAAAGGGGACCCCTCACTGTGTGTTAAAGAAAGAAGT	310
Qy	363	GGGAGAAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAAACAGCACACTG	422
Db	311	GGGAGAAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAAACAGCACACTG	370
Qy	423	AGATTAGGCACAGAGCAGACCATCTCTCTGTGTAAGTGACAAGCCACGCTCTTCTGGAAG	482
Db	371	AGATTAGGCACAGAGCAGACCATCTCTCTGTGTAAGTGACAAGCCACGCTCTTCTGGAAG	430
Qy	483	CCAAAGCTGACCAAGAGAAACAAATCCACCCAGACTCTAGACTCAGTCACTCTCTGAAG	542
Db	431	CCAAAGCTGACCAAGAGAGAAACAAATCCACCCAGACTCTAGACTCAGTCACTCTCTGAAG	490
Qy	543	CCCTCGTTTCAGGGTCGATATCCCCACATCAAGGACGGTGAGGATCTTTAAAGACCACTCAA	602
Db	491	CCCTCGTTTCAGGGTCGATATCCCCACATCAAGGACGGTGAGGATCTTTAAAGACCACTCAA	550
Qy	603	CAGAAAGTAAAAAAATGGAAAAATTTGCTAGAGAGAAATCCAGGCATGAAGTAGAAAAATCAG	662
Db	551	CAGAAAGTAAAAAAATGGAAAAATTTGCTAGAGAGAAATCCAGGCATGAAGTAGAAAAATCAG	610
Qy	663	AAATCACTGAAAAACACAGATGCTTCGGGCAAAAATAGAGAAATATAATGTTCCGCTGAACA	722
Db	611	AAATCACTGAAAAACACAGATGCTTCGGGCAAAAATAGAGAAATATAATGTTCCGCTGAACA	670
Qy	723	GGCTTAAGATGATGTTTGAGAAAGGTGAAACCAACTCAAACTAAGATTCTCCGGGCCCAAA	782
Db	671	GGCTTAAGATGATGTTTGAGAAAGGTGAAACCAACTCAAACTAAGATTCTCCGGGCCCAAA	730
Qy	783	GCCGAAGTCAAGTGGAGGAAGATCTCTCAAAACAGCTATTCTTAGATGACCTGGAAA	842
Db	731	GCCGAAGTCAAGTGGAGGAAGATCTCTCAAAACAGCTATTCTTAGATGACCTGGAAA	790
Qy	843	TAGGCCCCAGGTCAAGTTGTCATCTTCTACATTTGACTCGGAGAAAAATGAGAGTAGAGCA	902
Db	791	TAGGCCCCAGGTCAAGTTGTCATCTTTACATTTGACTCGGAGAAAAATGAGAGTAGAGCA	850
Qy	903	ATCTGAACTTCACGCCCTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAGG	962
Db	851	ATCTGAACTTCACGCCCTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAGG	910
Qy	963	CAGCTGTGTCAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGCGT	1022
Db	911	CAGCTGTGTCAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGCGT	970
Qy	1023	GGCAATCAAAATTCATAAAATGGAGCAAAAGAGAGATGTGCCCCAGGTCTCTGAGTCT	1082
Db	971	GGCAATCAAAATTCATAAAATGGAGCAAAAGAGAGATGTGCCCCAGGTCTCTGAGTCT	1030
Qy	1083	GCATCACCCATCAGGAAGGGGAAAAGATTCTGCAATGAGAAATAGCCTGGCAGTCGGTT	1142
Db	1031	GCATCACCCATCAGGAAGGGGAAAAGATTCTGCAATGAGAAATAGCCTGGCAGTCGGTT	1090
Qy	1143	CCACCCCTCCGAAGATGACTCCCGTGACTCCCGAGGTTAAGAGTGAGGTTCAACAGCGCTG	1202
Db	1091	CCACCCCTCCGAAGATGACTCCCGTGACTCCCGAGGTTAAGAGTGAGGTTCAACAGCGCTG	1150
Qy	1203	TCATCCCAAGCCACTAAGTCAGATTCAGAGACCTCCAGTCTTTCTGAAAGTTCTCCTC	1262

[illegible]

Db	2230	CTGAAGAATTCTACTACTCAGAATCAGAAATCCACGATGTGGAACCTCTCGGAGGGGAAG	2288
Qy	2343	TGGTCAAAGAGCTCTCTGTGGAAGAAGACAGATAAAGAGAAATCGGTATTATGATGAGGATG	2402
Db	2290	TGGTCAAAGAGCTCTCTGTGGAAGAAGACAGATAAAGAGAAATCGTATTATGATGAGGATG	2349
Qy	2403	AGGATGAAGAGTGCACAAATTGCAATGATGCTGGGCCCTTAAATTCATGTTAGTGTTAGCGA	2462
Db	2350	AGGATGAAGAGTGCACAAATTGCAATGATGCTGGGCCCTTAAATTCATGTTAGTGTTAGCGA	2409
Qy	2463	GCCACTGCCCTTTGTTCAAAATGTGATGCACATAGCAGGTATCCCGACATGAAATGTAAT	2522
Db	2410	GCCACTGCCCTTTGTTCAAAATGTGATGCACATAGCAGGTATCCCGACATGAAATGTAAT	2469
Qy	2523	TTACTTGGAAAGTAACTTTTGGAAAAGAATTCTTCTTTAAAAATCAAAA	2569
Db	2470	TTACTTGGAAAGTAACTTTTGGAAAAGAATTCTTCTTTAAAAATCAAAA	2516
RESULT 9			
AF157325			
LOCUS	AF157325	3348 bp mRNA linear PRI 04-MAY-2000	
DEFINITION	Homo sapiens sterol regulatory element binding protein 3 (SREBP3)		
ACCESSION	AF157325	mRNA, complete cds.	
VERSION	AF157325		
KEYWORDS	AF157325.1	GI:7688700	
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	2	(bases 1 to 3348)	
AUTHORS	Huang, Q., Gu, J., Ren, S., Jin, W., Gu, Y., Gu, W., Dong, H., Yu, Y., Fu, G., Wang, Y., Chen, Z. and Han, Z.		
TITLE	A novel gene expressed in the human hypothalamus		
JOURNAL	Unpublished		
REFERENCE	2	(bases 1 to 3348)	
AUTHORS	Ren, S., Shi, J., Huang, C., Jiang, C., Li, Y., Zhou, J., Yu, Y., Xu, S., Wang, Y., Fu, G., Chen, Z. and Han, Z.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-JUN-1999) Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China		
FEATURES	Location/Qualifiers		
source	1..3348		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/tissue_type="hypothalamus"		
gene	1..3348		
	/gene="SREBP3"		
CDS	691..2084		
	/gene="SREBP3"		
	/codon_start=1		
	/product="sterol regulatory element binding protein 3"		
	/protein_id="AAF67491.1"		
	/db_xref="GI:7688701"		
	/translation="MEQKNVPPGVEVCITHQGEKIKISANENSIAVRSTPAEDSDRS QVKEVQVHPKPSDPDSRSLSESSPPKMKKFPAPARETCVCKQTVVPMERLL ANQOVPH1SCFISYCNKLLSLGTYSALHGR1YCKPHNQLFKSGNTYDGFGRHPKH DLWAKNNEIEILERPAOLANARETPHSOGVEDAP1AKVGVLAASMEAKASQOEKEN RPAETTKLIAMPPTTELGSSGSALEEGIKSKPKWPPPEDE1SKPEVEDVLDLKLK RRSLSKRSRFRPTVAASFQTSVKSQPKTSPPIRKGWMSSEQSESVGGRVAERKQV ENAKASKNGNVGKTT1WQNKESKGTGKRSKEHGLEMENLNVENGADSDDDNSFL KQSPQEPKSLNWSFVDNFTFAEEFTTQNKSQDVELMEGEVVKLSLSEVEQIKRNRYY DEDEDEE"		
BASE COUNT	1052 a	620 c	703 g
ORIGIN			973 t

Query Match 62.7%; Score 2324; DB 9; Length 3348;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2694; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Figure 1. The effect of the number of trials on the number of correct responses. The number of correct responses was significantly higher for the 10 trials condition than for the 5 trials condition. Error bars represent the standard error of the mean.

QY 1001 AATGAGCTGAAGCCAGTGGTGGGAATCAAAATTCATAAATGGAGCAAAAGGAGAT 1060
Db 649 AATGAGCTGAAGCCAGTGGTGGGAATCAAAATTCATAAATGGAGCAAAAGGAGAT 708
QY 1061 GTGCCCCCAGGTCTCTGAGTCTGCATCAACCATCAGGAAGGGAAAGATTCTTGCAAAT 1120
Db 709 GTGCCCCCAGGTCTCTGAGTCTGCATCAACCATCAGGAAGGGAAAGATTCTTGCAAAT 768
QY 1121 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAGATGACTCCCGTGAATCCAGGTT 1180
Db 769 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAGATGACTCCCGTGAATCCAGGTT 828
QY 1181 AAGAGTGAAGTCAACAGCTCTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 1240
Db 829 AAGAGTGAAGTCAACAGCTCTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 888
QY 1241 AGTCTTTCTGAAGTCTCTCCCTCCCAAGCAATGAAGAATTTTCAGGCACCTGCAAGAGAG 1300
Db 889 AGTCTTTCTGAAGTCTCTCCCTCCCAAGCAATGAAGAATTTTCAGGCACCTGCAAGAGAG 948
QY 1301 ACTGGGTGAAGTGCAGAGACAGTCTATCCAAATGAGAGGTCTCTTGGCCAAACAGCAG 1360
Db 949 ACTGGGTGAAGTGCAGAGACAGTCTATCCAAATGAGAGGTCTCTTGGCCAAACAGCAG 1008
QY 1361 GTGTTTCACATCAGCTGCTTCCGTTGCTCTATTGCAACAACTCAGTCTAGGAACA 1420
Db 1009 GTGTTTCACATCAGCTGCTTCCGTTGCTCTATTGCAACAACTCAGTCTAGGAACA 1068
QY 1421 TATGCACTTTTACATGGAAGATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 1480
Db 1069 TATGCACTTTTACATGGAAGATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 1128
QY 1481 AAGGGCAACTATGATGAAGCTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAA 1540
Db 1129 AAGGGCAACTATGATGAAGCTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAA 1188
QY 1541 AATGAAACGAAGATTTTGGAGACAGACCCAGCTTGCAAATGCAAGGAGACCCCT 1600
Db 1189 AATGAAACGAAGATTTTGGAGACAGACCCAGCTTGCAAATGCAAGGAGACCCCT 1248
QY 1601 CACAGCCAGGGGTAGAAGATGCCCTATTGTAAGGTGGTCTGCTGCTCAAGTATG 1660
Db 1249 CACAGCCAGGGGTAGAAGATGCCCTATTGTAAGGTGGTCTGCTGCTCAAGTATG 1308
QY 1661 GAAGCCAAAGCCTCTCTCAGCAGGAGAGAGAGACAAGCCAGCTGAAACCAAGAGCTG 1720
Db 1309 GAAGCCAAAGCCTCTCTCAGCAGGAGAGAGAGACAAGCCAGCTGAAACCAAGAGCTG 1368
QY 1721 AGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTCAGGAAGTCCCTTGGAGAAAGG 1780
Db 1369 AGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTCAGGAAGTCCCTTGGAGAAAGG 1428
QY 1781 ATCAAAATGTCAAAGCCCAAAATGGCCCTCTGAAAGCAATCAGCAAGCCCAAGTCTCT 1840
Db 1429 ATCAAAATGTCAAAGCCCAAAATGGCCCTCTGAAAGCAATCAGCAAGCCCAAGTCTCT 1488
QY 1841 GAGGATGTCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGCAAAAGAGC 1900
Db 1489 GAGGATGTCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGCAAAAGAGC 1548
QY 1901 CGCCCATTCAGTGAAGCTTCACTTCAAGCACTCTGTCAAGAGCCCAAAACTGTG 1960
Db 1549 CGCCCATTCAGTGAAGCTTCACTTCAAGCACTCTGTCAAGAGCCCAAAACTGTG 1608
QY 1961 TCCCCACCTTATCAGGAAGGCTGGAGCATGTGAGAGCAGAGTGAAGTCTGTGGGTGGA 2020
Db 1609 TCCCCACCTTATCAGGAAGGCTGGAGCATGTGAGAGCAGAGTGAAGTCTGTGGGTGGA 1668
QY 2021 AGAGTTGCAAGAGAAACAAAGTGAAGTCCCAAGGCTTCTTGAAGCAATGGGATGTG 2080
Db 1669 AGAGTTGCAAGAGAAACAAAGTGAAGTCCCAAGGCTTCTTGAAGCAATGGGATGTG 1728
QY 2081 GGAAAAACAACTGGCAAAACAAAGATCTAAAGGAGAGACAGGGAAGTAAAGAA 2140

Db 1729 GGAAAAACAACTGGCAAAACAAAGATCTAAAGGAGAGACAGGGAAGAGATTAAGAA 1788
QY 2141 GGTCTAGTTCGGAGATGGAGAAATGAGAAATCTTTAGAAAAATGGTGCAGACTCCGATGAA 2200
Db 1789 GGTCTAGTTCGGAGATGGAGAAATGAGAAATCTTTAGAAAAATGGTGCAGACTCCGATGAA 1848
QY 2201 GATGATAACAGCTTCCTCAAAACAATCTCCACAAGAACCCAACTCTCTGAATTTGGTGG 2260
Db 1849 GATGATAACAGCTTCCTCAAAACAATCTCCACAAGAACCCAACTCTCTGAATTTGGTGG 1908
QY 2261 AGTTTTGTAGACAACACCTTTGCTGAAGAATTCACACTCAGAATCAGAAATCCAGGAT 2320
Db 1909 AGTTTTGTAGACAACACCTTTGCTGAAGAATTCACACTCAGAATCAGAAATCCAGGAT 1968
QY 2321 GTGGAACCTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGGAGAAACAGATAAGAGA 2380
Db 1969 GTGGAACCTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGGAGAAACAGATAAGAGA 2028
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAAATTTGCAATGATGCTGGCCCT 2440
Db 2029 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAAATTTGCAATGATGCTGGCCCT 2088
QY 2441 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCTTTGTCAAAATGTGATGCACATAAGCAG 2500
Db 2089 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCTTTGTCAAAATGTGATGCACATAAGCAG 2148
QY 2501 GTATCCCAAGCATGAATGTAATTTACTTTGGAAGTAACTTTGGAAAAAGAAATTCCTTCTTAA 2560
Db 2149 GTATCCCAAGCATGAATGTAATTTACTTTGGAAGTAACTTTGGAAAAAGAAATTCCTTCTTAA 2208
QY 2561 AATCAAAAAACAAACAAAAACAAAAACAAATTCATAATCTAGAGATACTTTTAC 2620
Db 2209 AATCAAAAAACAAACAAAAACAAAAACAAATTCATAATCTAGAGATACTTTTAC 2268
QY 2621 TTAATTTCTTCATTTTAGCAGTGATGATATGCATAAGTCTGTAAAGCTTCTTAAGCTGGG 2680
Db 2269 TTAATTTCTTCATTTTAGCAGTGATGATATGCATAAGTCTGTAAAGCTTCTTAAGCTGGG 2328
QY 2681 AAATATTCACCTGATTAATAGCCAGATTTCTGTAATCCCAAAAGSCAAATTTAAGGT 2740
Db 2329 AAATATTCACCTGATTAATAGCCAGATTTCTGTAATCCCAAAAGSCAAATTTAAGGT 2398
QY 2741 AGATAGATGATTAGTATATTTGTTACACACTATTTTGGAAATTAGAGAACATACAGAG 2800
Db 2399 AGATAGATGATTAGTATATTTGTTACACACTATTTTGGAAATTAGAGAACATACAGAG 2448
QY 2801 GAATTTAGGGGCTTTAAACATTTACGACTGAAATGCACTTTAGTATAAAGGGCACAGTTGTA 2860
Db 2449 GAATTTAGGGGCTTTAAACATTTACGACTGAAATGCACTTTAGTATAAAGGGCACAGTTGTA 2508
QY 2861 TATTTTAAATGAATACCAATTTAAATTTTAGTATTTTACCTGTTAAGAGATTTATTTAGT 2920
Db 2509 TATTTTAAATGAATACCAATTTAAATTTTAGTATTTTACCTGTTAAGAGATTTATTTAGT 2568
QY 2921 CTTTAAATTTTAGGTTAAATTTCTTGTGTGATATATATAGGAAATTTTACTATTTAT 2980
Db 2569 CTTTAAATTTTAGGTTAAATTTCTTGTGTGATATATATAGGAAATTTTACTATTTAT 2628
QY 2981 GTCTGTCTCTCAACCTACATCTCGACTCGAGCTGAGGTATATATACACAGAGCAC 3040
Db 2629 GTCTGTCTCTCAACCTACATCTCGACTCGAGGTATATATACACAGAGCAC 2688
QY 3041 TTTTGGAGCAATTTGAAAAACCAACCTTACACTCTTCGTCCTTAGAGAGATCTGCTGCT 3100
Db 2689 TTTTGGAGCAATTTGAAAAACCAACCTTACACTCTTCGTCCTTAGAGAGATCTGCTGCT 2748
QY 3101 CCCAAATAGGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTT 3160
Db 2749 CCCAAATAGGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTT 2808
QY 3161 TCTGGT-GATATCTGCTCTCTCATTAATTTACTGAAAGCTGCAATATTTTAGTAAATACCTT 3219

Db	2809	TCTGGTGATATCTGTCTCTCATTAATCTGAAAGCTGCAATATTTTAGTAATACCTT	2868
Qy	3220	CGGATACACTGCCCCCATCTTCGGTGTAGAGCAAGTGAAGTTTAAAGGAGGA	3279
Db	2869	CGGATCACTGCCCCCATCTTCGGTGTAGAGCAAGTGAAGTTTAAAGGAGGA	2928
Qy	3280	AGAAAGAACTGCTTTACACCACTTGAGCTCAGACCTCTAAACCTGTAATTCCTTATGA	3339
Db	2929	AGAAAGAACTGCTTTACACCACTTGAGCTCAGACCTCTAAACCTGTAATTCCTTATGA	2988
Qy	3340	TGTCCTCTTTTGAGACACTAAATTTTAAATCTACTAGCTCTGAAATATATGATTTT	3399
Db	2989	TGTCCTCTTTTGAGACACTAAATTTTAAATCTACTAGCTCTGAAATATATGATTTT	3048
Qy	3400	TATCACAGTATCTCAGGTGAATTTAAACCACTATAGGCTTTTCTTGGATGATTT	3459
Db	3049	TATCACAGTATCTCAGGTGAATTTAAACCACTATAGGCTTTTCTTGGATGATTT	3108
Qy	3460	TCTAGTCTTAAGTTTGGGACATATATAAATCTGAGTACATTTGTTGACACAGTTGATA	3519
Db	3109	TCTAGTCTTAAGTTTGGGACATATATAAATCTGAGTACATTTGTTGACACAGTTGATA	3168
Qy	3520	TTCCAAATTTGATGATGGGAGGAGGAGGTGCTTTAAGCTGTAGGCTTTCTTTGTAAGT	3579
Db	3169	TTCCAAATTTGATGATGGGAGGAGGAGGTGCTTTAAGCTGTAGGCTTTCTTTGTAAGT	3228
Qy	3580	CATTATAGAGATTTAGCTTTAATTTTATAGAGTGTAAACATCTCTCTTTCTTAGT	3639
Db	3229	CATTATAGAGATTTAGCTTTAATTTTATAGAGTGTAAACATCTCTCTTTCTTAGT	3288
Qy	3640	CTTACCTAGTCTGAACATTTTATTCATTAAGATTTTAAATTAATTTGAAAAA	3699
Db	3289	CTTACCTAGTCTGAACATTTTATTCATTAAGATTTTAAATTAATTTGAAAAA	3348
RESULT 10	AC008147/C		
LOCUS	AC008147	115345 bp	DNA linear PRI 29-MAR-2003
DEFINITION	Homo sapiens 12 BAC RP3-405J10 (Roswell Park Cancer Institute Human BAC Library) complete sequence.		
ACCESSION	AC008147		
VERSION	AC008147.31	GI:18464004	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 115345)		
AUTHORS	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayele, M., Banks, T., Barbara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Burch, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, F., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E.,		

Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S., Metker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pichens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., River, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G., Scher, S., Scott, G., Shen, H., Shim, C., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Zorrilla, S., Kuchelapatti, R., Weinstein, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 115345)
Worley, K.C.
Direct Submission
Submitted (27-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 115345)
Worley, K.C.
Direct Submission
Submitted (01-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 115345)
Worley, K.C.
Direct Submission
Submitted (28-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 115345)
Worley, K.C.
Direct Submission
Submitted (29-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Feb 1, 2002 this sequence version replaced gi:15626025.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

```
FEATURES          Location/Qualifiers
    source          1..115345
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="12"
                    /clone="RP3-405J10"
                    complement(2..2047)
                    /notes="overlaps bases 1..2020 of clone AC139016"
                    /function="clone overlap"
    repeat_region   complement(2..79)
    repeat_region   /rpt_family="AluY"
    repeat_region   110..156
    repeat_region   /rpt_family="(A)n"
    repeat_region   157..470
    repeat_region   /rpt_family="AluSg"
    repeat_region   471..484
    repeat_region   /rpt_family="(A)n"
    STS             998..1132
    STS             /standard_name="19465"
    STS             2114..2395
    repeat_region   /standard_name="133863"
    repeat_region   3365..3655
    repeat_region   /rpt_family="AluJb"
    repeat_region   3668..4037
    repeat_region   /rpt_family="LTR24"
    repeat_region   4041..4067
    repeat_region   /rpt_family="(TA)n"
    repeat_region   complement(4068..4242)
    repeat_region   /rpt_family="AluSg/X"
    repeat_region   complement(4243..4539)
    repeat_region   /rpt_family="AluSc"
    repeat_region   complement(4541..4836)
    repeat_region   /rpt_family="AluY"
    repeat_region   complement(4837..4890)
    repeat_region   /rpt_family="AluSg/X"
    repeat_region   complement(4891..5005)
    repeat_region   /rpt_family="AluJo"
    repeat_region   5006..5118
    repeat_region   /rpt_family="AluSx"
    repeat_region   5119..5425
    repeat_region   /rpt_family="AluSx"
    repeat_region   5426..5585
    repeat_region   /rpt_family="AluSx"
    repeat_region   complement(5586..5690)
    repeat_region   /rpt_family="AluJo"
    repeat_region   5931..6233
    repeat_region   /rpt_family="AluY"
    repeat_region   8022..8320
    repeat_region   /rpt_family="AluSg"
    repeat_region   8632..8934
    repeat_region   /rpt_family="AluSg"
    repeat_region   9461..9759
    repeat_region   /rpt_family="AluSx"
    repeat_region   9920..9947
    repeat_region   /rpt_family="(TTTTG)n"
    repeat_region   complement(9948..10229)
    repeat_region   /rpt_family="AluSg"
    repeat_region   11773..11943
    repeat_region   /rpt_family="LTR47A"
    repeat_region   complement(14835..15109)
    repeat_region   /rpt_family="AluSg"
    repeat_region   complement(15112..15412)
    repeat_region   /rpt_family="AluSg"
```

```
repeat_region      complement(15467..15779)
repeat_region      /rpt_family="AluSp"
repeat_region      15783..15818
repeat_region      /rpt_family="(T)n"
repeat_region      15932..16230
repeat_region      /rpt_family="AluJo"
repeat_region      16339..16637
repeat_region      /rpt_family="AluJb"
repeat_region      16640..16933
repeat_region      /rpt_family="AluY"
repeat_region      16934..16956
repeat_region      /rpt_family="(CAAA)n"
repeat_region      16962..17249
repeat_region      /rpt_family="AluSg"
repeat_region      17251..17288
repeat_region      /rpt_family="AT-rich"
repeat_region      18079..18112
repeat_region      /rpt_family="(TG)n"

Query Match      50.3%; Score 2234; DB 9; Length 115345;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 2284; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1408 CAGTCTAGGAACATATGCATCTTTACATGGAAGATCTATTGTGAAGCCTCACTTCAATCA 1467
DB 3172 CAGTCTAGGAACATATGCATCTTTACATGGAAGATCTATTGTGAAGCCTCACTTCAATCA 3113
QY 1468 ACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCT 1527
DB 3112 ACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCT 3053
QY 1528 ATGGCAAGCAAAAATGAAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTGCNAATGC 1587
DB 3052 ATGGCAAGCAAAAATGAAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTGCNAATGC 2993
QY 1588 AAGGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTATTGTAAGTGGGTGTCCT 1647
DB 2992 AAGGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTATTGTAAGTGGGTGTCCT 2933
QY 1648 GCCTGCAAGTATGAAGCCAGGCTCTCTCTCAGCAGGAGAGAGAACAGCAGCAGCTGA 1707
DB 2932 GCCTGCAAGTATGAAGCCAGGCTCTCTCTCAGCAGGAGAGAGAACAGCAGCAGCTGA 2873
QY 1708 AACCAAGAAGCTGAGGATCGCTCGGCCACCCCTGAACTTGAAGTTTCAGGAAGTGC 1767
DB 2872 AACCAAGAAGCTGAGGATCGCTCGGCCACCCCTGAACTTGAAGTTTCAGGAAGTGC 2813
QY 1768 CTTGGAGAGAGGATCAAAAATGTCAAAAGCCCAATGGCCTCTGAAAGCGAAATCAGCAA 1827
DB 2812 CTTGGAGAGAGGATCAAAAATGTCAAAAGCCCAATGGCCTCTGAAAGCGAAATCAGCAA 2753
QY 1828 GCCGGAAGTTCTTGAGGATGCGATCTAGATCTGAAGAAGCTAAGACGATCTTCTCACT 1887
DB 2752 GCCGGAAGTTCTTGAGGATGCGATCTAGATCTGAAGAAGCTAAGACGATCTTCTCACT 2693
QY 1888 GAAGGAAGAAGCGCCCATTTCACTGTAGCAGGCTTCATTTCAAGACCTCTGTCAAGAG 1947
DB 2692 GAAGGAAGAAGCGCCCATTTCACTGTAGCAGGCTTCATTTCAAGACCTCTGTCAAGAG 2633
QY 1948 CCCAAAAAAGTGTGTCCCACTATTCAGGAAGGCTGGAGCATGTGAGAGAGAGTGAAGA 2007
DB 2632 CCCAAAAAAGTGTGTCCCACTATTCAGGAAGGCTGGAGCATGTGAGAGAGAGTGAAGA 2573
QY 2008 GTCTGTGGGTGGAAGAGTTGCAAAAAGGAAACAAAGTGGAAAAATGCCAAGGCTTCAAGAA 2067
DB 2572 GTCTGTGGGTGGAAGAGTTGCAAAAAGGAAACAAAGTGGAAAAATGCCAAGGCTTCAAGAA 2513
QY 2068 GAATGGGAATGTGGGAACACCTGGCAAAACAAAGATCTAAAGGAGAGACAGGAA 2127
DB 2512 GAATGGGAATGTGGGAACACCTGGCAAAACAAAGATCTAAAGGAGAGACAGGAA 2453
QY 2128 GAGAAGTAAAGAGGTCATAGTTTGGAGATGGAGATGAGATCTTTAGAAAAATGGTGC 2187
DB 2188 GAGAAGTAAAGAGGTCATAGTTTGGAGATGGAGATGAGATCTTTAGAAAAATGGTGC 2187
```

Db	2452	GAGAAGTAAGGAAGGTCATAGTTTGGAGATGGAGAATGAGAATCTTGAGAAAAATGGTGC	2399
QY	2188	AGACTCCGATGAAGATGATAACAGCTTCCTCAACACACAATCTCCACAAGAACCCAGTCT	2247
Db	2392	AGACTCCGATGAAGATGATAACAGCTTCCTCAACACAATCTCCACAAGAACCCAGTCT	2333
QY	2248	TCTGAATTGGTCAGGTTTTCTAGACAAACACCTTTGCTGAAGAATTCCTACTACTCAGAATCA	2307
Db	2332	TCCTGAATTGGTCAGGTTTTCTAGACAAACACCTTTGCTGAAGAATTCCTACTACTCAGAATCA	2273
QY	2308	GAATCCAGATGTGGAACTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGTGAAGA	2367
Db	2272	GAATCCAGATGTGGAACTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGTGAAGA	2213
QY	2368	ACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGCACAAATTCGAAT	2427
Db	2212	ACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGCACAAATTCGAAT	2153
QY	2428	GATGCTGGGCTTTAAATTCATGTTAGTGTGTAGCGAGCCACTGCCCTTTGTCAAAAATGTGA	2487
Db	2152	GATGCTGGGCTTTAAATTCATGTTAGTGTGTAGCGAGCCACTGCCCTTTGTCAAAAATGTGA	2093
QY	2488	TGCACATAAGCAGGTATCCAGCATGAAATGTAATTTACTTTGGAAGTAACTTTGGAAAAG	2547
Db	2092	TGCACATAAGCAGGTATCCAGCATGAAATGTAATTTACTTTGGAAGTAACTTTGGAAAAG	2033
QY	2548	AATTCCTTCCTAAAAATCAAAAAACAAAAACAAAAACAAAAACAACTCTAAATACTA	2607
Db	2032	AATTCCTTCCTAAAAATCAAAAAACAAAAACAAAAACAAAAACAACTCTAAATACTA	1973
QY	2608	GAGATAACTTTACTTAAATTCCTCATTTTAGCAGTGATGATATGCATAAGTGTCTGAAGG	2667
Db	1972	GAGATAACTTTACTTAAATTCCTCATTTTAGCAGTGATGATATGCCTAAGTGTCTGAAGG	1913
QY	2668	CTTGTAACCTGGGGAATATTCACACCTGATATATAGCCAGATTCCTACTGTATTCCTCAAAAG	2727
Db	1912	CTTGTAACCTGGGGAATATTCACACCTGATATATAGCCAGATTCCTACTGTATTCCTCAAAAG	1853
QY	2728	GCAATATTAAGGTAGATAGATGATTAGTAGTATATTTGTTACACACTATTTTGGAAATAGA	2787
Db	1852	GCAATATTAAGGTAGATAGATGATTAGTAGTATATTTGTTACACACTATTTTGGAAATAGA	1793
QY	2788	GAACATACAGAAGGAATTTAGGGGCTTAAACATTAAGCTGAATGCACTTAGTATAAAG	2847
Db	1792	GAACATACAGAAGGAATTTAGGGGCTTAAACATTAAGCTGAATGCACTTAGTATAAAG	1733
QY	2848	GGCACAGTTTGATATATTTAAATGAATACCAATTTAAATTTTAGTATTTACCTGTTAA	2907
Db	1732	GGCACAGTTTGATATATTTAAATGAATACCAATTTAAATTTTAGTATTTACCTGTTAA	1673
QY	2908	GAGATTATTAGTCTTTAAATTTTTAGGTTAAATTTCTGCTGCATATATATAGAGAA	2967
Db	1672	GAGATTATTAGTCTTTAAATTTTTAGGTTAAATTTCTGCTGCATATATATAGAGAA	1613
QY	2968	TTTACTACTTTATGCTGCTCTCTAAATTAACATCTGAATCGACGCTCTGAGGTATAA	3027
Db	1612	TTTACTACTTTATGCTGCTCTCTCTAAATTAACATCTGAATCGACGCTCTGAGGTATAA	1553
QY	3028	TACAACAGAGCACTTTTGGGCAATTTGAARACCAACCTACACTCTTCGGTGCTTAGAG	3087
Db	1552	TACAACAGAGCACTTTTGGGCAATTTGAARACCAACCTACACTCTTCGGTGCTTAGAG	1493
QY	3088	AGATCTGCTGTCTCCAAAATAAGCTTTGTATCTGCCAGTGAATTTTACTGTACTCCAAAT	3147
Db	1492	AGATCTGCTGTCTCCAAAATAAGCTTTGTATCTGCCAGTGAATTTTACTGTACTCCAAAT	1433
QY	3148	GATTGCTTTCTTTTCTGGTGATATCTGTGCTTCTCAATAATTTACTGAAAGCTGCAATTTT	3207
Db	1432	GATTGCTTTCTTTTCTGGTGATATCTGTGCTTCTCAATAATTTACTGAAAGCTGCAATTTT	1373
QY	3208	TAGTAATACTCTCGGATACACTGTGCCCATCTCTCGTCTGTAGAGCAAGTGAAGTTTT	3267
Db	1372	TAGTAATACTCTCGGATACACTGTGCCCATCTCTCGTCTGTAGAGCAAGTGAAGTTTT	1313

Qy	3268	AAAGGAGGAAGAAGAAGAACTCTCTTTACACCACTTTGAGCTCAGACCTCTAAACCCCTGTA	3327
Db	1312	AAAGGAGGAAGAAGAAGAACTCTCTTTACACCACTTTGAGCTCAGACCTCTAAACCCCTGTA	1253
Qy	3328	TTTCCCTTATGATGTCCTCCCTTTTGTAGAGACACTAAATTTTAAATTACTTACTAGCTCTGAAA	3387
Db	1252	TTTCCCTTATGATGTCCTCCCTTTTGTAGAGACACTAAATTTTAAATTACTTACTAGCTCTGAAA	1193
Qy	3388	TATATTGATTTTATCACAGTATCTTCAGGGTGAAATTAACCAACTATAGGCCCTTTTTC	3447
Db	1192	TATATTGATTTTATCACAGTATCTTCAGGGTGAAATTAACCAACTATAGGCCCTTTTTC	1133
Qy	3448	TTGGGATGATTTTCTTAGTCTTTAGGTTTGGGACATTAATAAACTTCAGTACATTTGTTGT	3507
Db	1132	TTGGGATGATTTTCTTAGTCTTTAGGTTTGGGACATTAATAAACTTCAGTACATTTGTTGT	1073
Qy	3508	ACACAGTTGATTTCCAAATTTGATCGATGGGAGGAGGTGTCCTTAAGCTGTAGGCTT	3567
Db	1072	ACACAGTTGATTTCCAAATTTGATCGATGGGAGGAGGTGTCCTTAAGCTGTAGGCTT	1013
Qy	3568	TTCTTTGTACTGCATTTATAGAGATTTAGCTTTAATATTTTTTAGAGATGTAAGACATTC	3627
Db	1012	TTCTTTGTACTGCATTTATAGAGATTTAGCTTTAATATTTTTTAGAGATGTAAGACATTC	953
Qy	3628	TGCTTTCTTAGTCTTACCTAGTCTGAAAATTTTATTCAAATAAGATTTAATTAATAAT	3687
Db	952	TGCTTTCTTAGTCTTACCTAGTCTGAAAATTTTATTCAAATAAGATTTAATTAATAAT	893
Qy	3688	TTGAA 3692	
Db	892	TTGAA 888	

RESULT 11

BC010664

LOCUS ..

DEFINITION Homo sapiens, clone IMAGE:3854371, mRNA, partial cds.

2121 bp mRNA linear PRI 12-JUL-2001

RESULT 11
BC010664

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

REFERENCE
AUTHORS
TITLE
JOURNAL.

REMARK
COMMENT

BC010664 2121 bp mRNA linear PRI 12-JUL-2001
Homo sapiens, clone IMAGE:3854371, mRNA, partial cds.
BC010664
BC010664.1 GI:14715008

. Homo sapiens (human)

Homo sapiens

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2121)

Strausberg, R.

Direct Submission

Submitted (10-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590.

USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-rc@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.basc.bcm>

Web site: <http://www.ngsc.bc.ca/cmc.edu/cana/>
Contact: villalon@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 13 Row: b Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES	source	Location/Qualifiers
		1. .2121
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone="IMAGE:3854371"
		/tissue_type="Colon, adenocarcinoma"
		/clone_lib="NIH MGC_65"
		/lab_host="DH10B"
		/note="Vector: pCMV-SPORT6"
CDS		<1..831
		/codon_start=1
		/product="unknown (protein for IMAGE:3854371)"
		/protein_id="AAH10664.1"
		/db_xref="GI:14715009"
		/translation="ARETSPHSGVEDAPIAKVGLAASMEAKASSOQEKEDKPAETTK LRIANPPPELGSAGLBEIGIKMSKPKPPEDEIKSPVEDVDLDLKKLRSSSLK ERRPFTAAASFQSTSVKSPKTPVPIRKGMSNSESESVGRVAERKQVENAKASK KNGNKVTKWQNKESGTGKRSKEGHSLEMENILVENGADSDSDNDFLKKQSPQE PKSLNNSFSFYDNTFAEFTTQNKQSDVELWEQVVKELSVBEQIKRNRYYDEDEDE "
BASE COUNT	697 a	376 c 448 g 600 t
ORIGIN		
		Query Match 53.1%; Score 1967; DB 9; Length 2121;
		Best Local Similarity 99.9%; Pred. No. 0;
		Matches 2117; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	1586	GCAAGGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTTATGCTAAGTGGGTGTC 1645
Db	1	GCAAGGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTTATGCTAAGTGGGTGTC 60
QY	1646	CTGGCTCAAGTATGGAAGCCAAAGGCCTCCTCTCAGCAGGAGGAAGACAAGCCAGCT 1705
Db	61	CTGGCTCAAGTATGGAAGCAAGGCCTCCTCTCAGCAGGAGGAAGACAAGCCAGCT 120
QY	1706	GAACCAAGAAGCTGAGGATCGGCTGCCACCCGCCACTGAACTTGAAGTTCAGGAAGT 1765
Db	121	GAACCAAGAAGCTGAGGATCGGCTGCCACCCGCCACTGAACTTGAAGTTCAGGAAGT 180
QY	1766	GCCTTGAGGAAGGGATCAAAATGTCAAGCCCAATGGCCTCTGAAAGCAAGAAATCAGC 1825
Db	181	GCCTTGAGGAAGGGATCAAAATGTCAAGCCCAATGGCCTCTGAAAGCAAGAAATCAGC 240
QY	1826	AAGCCCGAAGTCTCTGAGGATGTGATCTAGATCTGAAGAAGCTAGACGATCTTCTTCA 1885
Db	241	AAGCCCGAAGTCTCTGAGGATGTGATCTAGATCTGAAGAAGCTAGACGATCTTCTTCA 300
QY	1886	CTGAAGAAAGAAAGCCGCCCATTCAGTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAG 1945
Db	301	CTGAAGAAAGAAAGCCGCCCATTCAGTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAG 360
QY	1946	AGCCCAAAACTGTGCCCCACCTATCAGAAAGGCTGGAGCATGTCAGAGCAGAGTGAA 2005
Db	361	AGCCCAAAACTGTGCCCCACCTATCAGAAAGGCTGGAGCATGTCAGAGCAGAGTGAA 420
QY	2006	GAGTCTGTGGTGGAGAGTTCAGAAAGGAAAACAAGTGGAAAATGCCAAGGCTTCTTAAG 2065
Db	421	GAGTCTGTGGTGGAGAGTTCAGAAAGGAAAACAAGTGGAAAATGCCAAGGCTTCTTAAG 480
QY	2066	AAGAATGGGAATGTGGGAAAAAACAACCTGCCAAAAACAAGAATCTAAGAGAGACAGGG 2125
Db	481	AAGAATGGGAATGTGGGAAAAAACAACCTGCCAAAAACAAGAATCTAAGAGAGACAGGG 540
QY	2126	AAGAGAAGTAAGCAAGGTCTAGCTTTGGAGATGGAGATGAGAAATCTTGTAGAAAATGGT 2185
Db	541	AAGAGAAGTAAGCAAGGTCTAGCTTTGGAGATGGAGATGAGAAATCTTGTAGAAAATGGT 600
QY	2186	GCAGATCTCCGATGAAGATGATAACAGCTTCCTCAAAACAACAATCTCCACAAGAACCCAAG 2245
Db	601	GCAGATCTCCGATGAAGATGATAACAGCTTCCTCAAAACAACAATCTCCACAAGAACCCAAG 660
QY	2246	TCTCTGAATGGTCGATTTTGTAGACAAACACCTTTGCTGAAGAAATCTACTACTCAGAAAT 2305

Db 1741 TATTTCCTTATGATGTCCTCTTTTGAGACACTAAATTTTAAATCTACTAGCTCTGA 1800

Qy 3386 AATATATTGATTTTATACAGTATTTCTCAGGGTAAATTAACCACTATAGCCCTTTT 3445

Db 1801 AATATATTGATTTTATACAGTATTTCTCAGGGTAAATTAACCACTATAGCCCTTTT 1860

Qy 3446 TCTTGGGATGATTTTCTAGTCTTAAAGTTTGGGACATTAATACTTGGATGATTTGTT 3505

Db 1861 TCTTGGGATGATTTTCTAGTCTTAAACGTTTGGGACATTAATACTTGGATGATTTGTT 1920

Qy 3506 GTACACAGTTCATATTCACAAATTTGATGGATGGGAGGAGGTGCTTAAGCTGTAGGC 3565

Db 1921 GTACACAGTTCATATTCACAAATTTGATGGATGGGAGGAGGTGCTTAAGCTGTAGGC 1980

Qy 3566 TTTCTTTTGTACTGATTTATAGAGATTTAGCTTTAAATTTTATAGAGATGTAACACAT 3625

Db 1981 TTTCTTTTGTACTGATTTATAGAGATTTAGCTTTAAATTTTATAGAGATGTAACACAT 2040

Qy 3626 TCTGCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATCAATAAGATTTAAATTTAA 3685

Db 2041 TCTGCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATCAATAAGATTTTAAATTTAA 2100

Qy 3686 ATTTGAAAAAATAAAAAA 3705

Db 2101 ATTTGAAAAAATAAAAAA 2120

RESULT 12

AK000057 2667 bp mRNA linear PRI 22-FEB-2000

LOCUS Homo sapiens cDNA FLJ20050 fis, clone COL00688.

DEFINITION AK000057

ACCESSION AK000057.1 Gi:7019894

VERSION oligo capping; fis (full insert sequence).

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (sites)

AUTHORS Kawabata,A., Hiki,J., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.

TITLE NEDO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2667)

AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.

TITLE Direct Submission

JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES

source

1. .2667

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="COL00688"

/tissue_type="colon"

/clone_lib="COL"

/note="cloning vector pME18SFL3"

10. .1380

/note="unnamed protein product"

/codon_start=1

/protein_id="BAA30914.1"

/db_xref="GI:7019895"

CDS

/translation="MEQENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSDRS QYKSEVQHPKPLSPUSRASSSESSPPKAMFQAPARETCVSCQKTVTPMERLLA NOQVHISCFRCSCYNKLSLGYASLRGRIYCKPHFNQLFKSKNGYDEGFQHPHK LWAASNEBEILERPAQLANARETPHSPGVADIAKGVLAASMEAKASQOEKEDK PAETKRLRTAMPPTPELSSGSALEEGIKMSKPWPPEDEISKPEVPEDVDLKLRL RSSLSKERSRPETVAASQSTSVKSPKTVSPPIRKMSMSBQSESVESVQVUARQVE NAKASKONGNVKTTWKNKSKGRTGKSKSKEGHSLEMEMENLVENGASDEDDNSFLK QSQPQPSKLSNWSFVDTTFAEFTTQNKQSDVELMEGVEVVKELSEVBEQIKRNYD EDEDEE"

BASE COUNT 865 a 512 c 570 g 720 t

ORIGIN

Query Match 51.4%; Score 1903; DB 9; Length 2667;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1953; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1276 GAAGTTTCAGGCACCTGCGAAGAGAGACCTGCGTGGAAATGTCAGAGACAGTCTATCCCAAT 1335

Db 240 GAAGTTTCAGGCACCTGCGAAGAGAGACCTGCGTGGAAATGTCAGAGACAGTCTATCCCAAT 299

Qy 1336 GGAGCGTCTCTTGGGCAACACGACAGGTGTTTCACATCAGCTGCTTCCGTTGCTCTATTG 1395

Db 300 GGAGCGTCTCTTGGGCAACACGACAGGTGTTTCACATCAGCTGCTTCCGTTGCTCTATTG 359

Qy 1396 CAACAAACAACTCAGTCTAGGAACATATGCATCTTTTACATGGAAGAATCTATTGTAAGCC 1455

Db 360 CAACAAACAACTCAGTCTAGGAACATATGCATCTTTTACATGGAAGAATCTATTGTAAGCC 419

Qy 1456 TCATCTCAATCAACTCTTTTAAATCTTAAGGCAACTATGATGAAGGCTTTGGGACAGACC 1515

Db 420 TCATCTCAATCAACTCTTTTAAATCTTAAGGCAACTATGATGAAGGCTTTGGGACAGACC 479

Qy 1516 ACACAGGATCTATGGGCAAGCAAAAATGAAAACGAAGAGATTTTGGAGAGACAGCCCA 1575

Db 480 ACACAGGATCTATGGGCAAGCAAAAATGAAAACGAAGAGATTTTGGAGAGACAGCCCA 539

Qy 1576 GCTTGCATAATGCAAGGGAGACCCCTCACAGCCAGCGGTAGGAAGATGCCCTTATTGCTAA 1635

Db 540 GCTTGCATAATGCAAGGGAGACCCCTCACAGCCAGCGGTAGGAAGATGCCCTTATTGCTAA 599

Qy 1636 GTGGGTGTCCTGGTGCAGATGATGGAAGCAAGGCTCTCTCAGCAGGAGGAAGAAGA 1695

Db 600 GTGGGTGTCCTGGTGCAGATGATGGAAGCAAGGCTCTCTCAGCAGGAGGAAGAAGA 659

Qy 1696 CAAGCAGCTGAAACCAAGAGCTGAGGATGCGCTGGCCACCCCTGACCTGGAAG 1755

Db 660 CAAGCAGCTGAAACCAAGAGCTGAGGATGCGCTGGCCACCCCTGACCTGGAAG 719

Qy 1756 TTCAGGAAGTGCCTTGGAGAGAGGATCAAAATGTCAAAGCCCAATGGCCTCTTGAAGA 1815

Db 720 TTCAGGAAGTGCCTTGGAGAGAGGATCAAAATGTCAAAGCCCAATGGCCTCTTGAAGA 779

Qy 1816 CGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTGAAGAAGCTAAGACG 1875

Db 780 CGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTGAAGAAGCTAAGACG 839

Qy 1876 ATCTTCTTCACTGAAGGAAGAAGCGGCCCATTCATCTGAGCAGTCTTCAATTTCAAGCAC 1935

Db 840 ATCTTCTTCACTGAAGGAAGAAGCGGCCCATTCATCTGAGCAGTCTTCAATTTCAAGCAC 899

Qy 1936 CTCTGTCAAGGCCCAAAACCTGTCTCCCACTATCAGGAAGGCTGGAGATCTCAGA 1995

Db 900 CTCTGTCAAGGCCCAAAACCTGTCTCCCACTATCAGGAAGGCTGGAGATCTCAGA 959

Qy 1996 GCAGAGTGAAGAGTCTGTGGGTGGAAAGATTCAGAAAGGAACCAAGTGGAAATTCGCAA 2055

Db 960 GCAGAGTGAAGAGTCTGTGGGTGGAAAGATTCAGAAAGGAACCAAGTGGAAATTCGCAA 1019

Qy 2056 GGCCTTCTAAGAAAGATGGAATGTGGGAAAAACAACTGGGAAAAACAAAGATCTTAAAGG 2115

Db 1020 GGCCTTCTAAGAAAGATGGAATGTGGGAAAAACAACTGGGAAAAACAAAGATCTTAAAGG 1079

Qy 2116 AGACAGCGGAGAGAAAGTAAGAGGTCATAGTTTGGAGATGGGAATGAGATCTTGT 2175

1080	AGACAGGGAAGAGAGATGAAGAGGT	CATAGTTTGGAGATGGAGATGAGAACTCTTGT	1139
2176	AGAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCCTCAAACCAACAACTCTCCACA	2235	
1140	AGAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCCTCAAACCAACAACTCTCCACA	1199	
2236	AGAACCCAGTCTCTGAAATTTGGTTCGAGTTTGTAGACAAACACCTTTGCTGAAGAAATTCAC	2295	
1200	AGAACCCAGTCTCTGAAATTTGGTTCGAGTTTGTAGACAAACACCTTTGCTGAAGAAATTCAC	1259	
2296	TACTCAGAAATCAGAAATCCCAGGATGTGGAACCTCTGGGAGGGAAGTGGTCAAAGCT	2355	
1260	TACTCAGAAATCAGAAATCCCAGGATGTGGAACCTCTGGGAGGGAAGTGGTCAAAGCT	1319	
2356	CTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGGAAGTG	2415	
1320	CTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGGAAGTG	1379	
2416	ACAAATTCGAATGATGCTGGGCTTTAAATTCATGTTAGTCTTAGCGAGGCACCTGCCCTTT	2475	
1380	ACAAATTCGAATGATGCTGGGCTTTAAATTCATGTTAGTCTTAGCGAGGCACCTGCCCTTT	1439	
2476	GTCAAAATGTGATGCACATTAAGCAGGTATCCGACGATGAAATGTAATTTACTTTGGAAGTA	2535	
1440	GTCAAAATGTGATGCACATTAAGCAGGTATCCGACGATGAAATGTAATTTACTTTGGAAGTA	1499	
2536	ACTTTGGAAAGAAATTCCTTCTTTAAATCAAACCAAAACAAAAACACAAAAACACA	2595	
1500	ACTTTGGAAAGAAATTCCTTCTTTAAATCAAACCAAAACAAAAACACAAAAACACA	1559	
2596	TTCTAAATACTAGAGATAACTTTACTTAAATCTCTTATTTAGCAGTGATGATATGCATA	2655	
1560	TTCTAAATACTAGAGATAACTTTACTTAAATCTCTTATTTAGCAGTGATGATATGCATA	1619	
2656	AGTGTGTAGGCTTGTAACTGGGAAATATTCCA	2715	
1620	AGTGTGTAGGCTTGTAACTGGGAAATATTCCA	1679	
2716	TATTTCCAAAGGCAATATTAAAGGTAGATGATTAGTAGTATATTTGTTACACACTAT	2775	
1680	TATTTCCAAAGGCAATATTAAAGGTAGATGATTAGTAGTATATTTGTTACACACTAT	1739	
2776	TTTGGAAATTAGAAACATACAGAGGAATTTAGGGCTTAAACATTACGACTGAATGCAAC	2835	
1740	TTTGGAAATTAGAAACATACAGAGGAATTTAGGGCTTAAACATTACGACTGAATGCAAC	1799	
2836	TTTAGTATAAAGGCGACAGTTTCTATATTTTAAATGAATACCAATTTAATTTTTTAGTA	2895	
1800	TTTAGTATAAAGGCGACAGTTTCTATATTTTAAATGAATACCAATTTAATTTTTTAGTA	1859	
2896	TTTACCTGTTTAAGAGATTATTAGTCTTTAAATTTTATAGGTTAATTTTCTGCTGTGAT	2955	
1860	TTTACCTGTTTAAGAGATTATTAGTCTTTAAATTTTATAGGTTAATTTTCTGCTGTGAT	1919	
2956	ATATATGAGAAATTTACTACTTTATGCTGCTCTCTAAACTACATCCTGAACTGCACGT	3015	
1920	ATATATGAGAAATTTACTACTTTATGCTGCTCTCTAAACTACATCCTGAACTGCACGT	1979	
3016	CCTGAGGTATATACAAACAGAGCACTTTTGGAGCAATTTGAAACCAACCTACACTCTT	3075	
1980	CCTGAGGTATATACAAACAGAGCACTTTTGGAGCAATTTGAAACCAACCTACACTCTT	2039	
3076	CGGTGCTTAGAGAGATCTGCTGCTCCCAAAATGAGCTTTTGATCTGCCAGTGAATTTAC	3135	
2040	CGGTGCTTAGAGAGATCTGCTGCTCCCAAAATGAGCTTTTGATCTGCCAGTGAATTTAC	2099	
3136	TGTACTCCAAATGATTGCTTTCTTTTCTGGTGATATCTGTGTTCTCAATAATTTACTGAAA	3195	
2100	TGTACTCCAAATGATTGCTTTCTTTTCTGGTGATATCTGTGTTCTCAATAATTTACTGAAA	2159	
3196	GCTGCAATATTTTGTAGTAATACCTTCGGATCACT	3229	

Db	2160	GTGGCAATATTTAGTAATACCTTCGGATCACT	2193
RESULT 13			
BC001247			
LOCUS	BC001247	2164 bp	mrna linear PRI 12-JUL-2001
DEFINITION	Homo sapiens, epithelial protein lost in neoplasm beta, clone		
	MGC:4969 IMAGE:3452714, mRNA, complete cds.		
ACCESSION	BC001247		
VERSION	BC001247.1	GI:12654808	
KEYWORDS	MGC		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 2164)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-DEC-2000) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-3590,		
	USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk		
	Email: cgabbs-remail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Sequencing Group at the Stanford Human Genome		
	Center, Stanford University School of Medicine, Stanford, CA 94305		
	Web site: http://www.shgc.stanford.edu		
	Contact: (Dickson, Mark) mcd@paxil.stanford.edu		
	Dickson, M., Schmutz, J., Grinstead, J., Rodriguez, A., and Myers,		
	R. M.		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E.B. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 4 Row: 0 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA q1: 7020350.

FEATURES
source

```

Location/Qualifiers
1. 2164
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:51474"
/db_xref="taxon:9606"
/clones="MGC:4969 IMAGE:3452714"
/tissue_type="Placenta, Chorionic
Tissue_Type="M11 MGC 10"
/lab_host="DH10B"
/notes="vector: pCMV-SPORT6"
187. 1983

```

CDS

```
/lab_host="DHIUB"  
/note="Vector: pCMV-SPORT6"  
187. 1983
```

```

/codon_start=1
/product="epithelial protein lost in neoplasm beta"
/protein_id="AAH01247.1"
/db_xref="GI:12654809"

```

/translation= "WENCLGESRHEVEKSEISENTDASGKIEKYNVPLNRLXOMFEKG
EPTQIKLRASGRAGKRSSENYSLDDUEJGPGUUSZTDFSEKNESRNLNRLP
SETSIKORMAKQAASQKSSNTYLLKASGEIJIHKWENQVPPGPEVCIITH
EGEKSIAENSLVETPTADSDSDSQVQOPKPLSPDRASLSSESPPKA
MKFOAPARETVCQCKTQYPMERLANQVFIHSFPCFSCYNNKLUGTYSALHGRY
CKPHFNOLFKSNGNDGPHRKHDLWASNEVLEILERPALANARETHSPGVED
APIALGVGLASMEAKASQOEKDEPAETKPIIAMPPTLGGSSCALZEGIKMSK
PKWPEDEJISKPEYEDVDLJLKLRSLSLKERSRPFVVAASFTVSQSPKTVSP
IRKWSMNSGSESGBRVAORNAKAKNGNNGVTKTQYNQKBSKGTGKRSKG
HSLMEBNLENVAGSDSDNSLQKQSQEPKQSLWMSFVDNTFAEETQYNQKS
DVLWEJGSEVKESEVBEQIKRNYRYSDEDEE"

[illegible]

DELWEGEVKELSVEEQIKRNRYYDE
a 472 C 513 g 435 t

Query Match

51.3%; Score 1901; DB 9; Length 2164;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 2141; Conservative 0; Mismatches

it

QY 431 CACAGAGCAGACCATCTCTCTGTGTAAGTGACAAAGCCACGCTGTTCTGAGGCCAAAGCT 490
Db 1 CACAGAGCAGACCATCTCTCTGTGTAAGTGACAAAGCCACGCTGTTCTGAGGCCAAAGCT 60
QY 491 GACCAAGAGAACAAATCCACCCAGATCTAGACTCAGGTCACCTCTGAGCCCTCGTT 550
Db 61 GACCAAGAGAACAAATCCACCCAGATCTAGACTCAGGTCACCTCTGAGCCCTCGTT 120
QY 551 CAGGTCGATATCCCCACATCAAGGACGGTGAGGATCTTAAAGACCACTCAACAGAAAGT 610
Db 121 CAGGTCGATATCCCCACATCAAGGACGGTGAGGATCTTAAAGACCACTCAACAGAAAGT 180
QY 611 AAAAAATGGAATTTGTCTAGGAGAAATCCAGGCATGAAGTAGAAAAATCAGAAATCAGT 670
Db 181 AAAAAATGGAATTTGTCTAGGAGAAATCCAGGCATGAAGTAGAAAAATCAGAAATCAGT 240
QY 671 GAAACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAG 730
Db 241 GAAACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAG 300
QY 731 ATGATGTTTGAGAAAGGTGAACCACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGT 790
Db 301 ATGATGTTTGAGAAAGGTGAACCACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGT 360
QY 791 GCAAGTGAAGAGAGATCTCTGAAAAAGAGCTATTCTTAGATGACCTGGAAATAGGCCCA 850
Db 361 GCAAGTGAAGAGAGATCTCTGAAAAAGAGCTATTCTTAGATGACCTGGAAATAGGCCCA 420
QY 851 GGTGAGTTGTCATCTTACATTTGACTCGGAGAAAAATGAGAGTAGCAAAATCTGGAA 910
Db 421 GGTGAGTTGTCATCTTACATTTGACTCGGAGAAAAATGAGAGTAGCAAAATCTGGAA 480
QY 911 CTTCCAGCCTCTCAGAAACCTCTATAAAGGATCGAATGCGCAAGTACCAAGGAGCTGTG 970
Db 481 CTTCCAGCCTCTCAGAAACCTCTATAAAGGATCGAATGCGCAAGTACCAAGGAGCTGTG 540
QY 971 TCCAAACAAAGACGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGTGCGCAATC 1030
Db 541 TCCAAACAAAGACGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGTGCGCAATC 600
QY 1031 AAAATTCTAAATTTGGAGCAAAAGAGAAATGTCGCCAGGTCCTGAGGTCCTGATCACC 1090
Db 601 AAAATTCTAAATTTGGAGCAAAAGAGAAATGTCGCCAGGTCCTGAGGTCCTGATCACC 660
QY 1091 CATCAGGAGGGGAAAGATTCTCGAAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCT 1150
Db 661 CATCAGGAGGGGAAAGATTCTCGAAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCT 720
QY 1151 GCCGAAGATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCTGTCCATCCC 1210
Db 721 GCCGAAGATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCTGTCCATCCC 780
QY 1211 AAGCCACTAAGTCCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTCTCTCCCAAGCA 1270
Db 781 AAGCCACTAAGTCCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTCTCTCCCAAGCA 840
QY 1271 ATGAAGAGTTTCAGGACCTGCAAGAGAGACCTGCGTGGNATGTGAGAGACAGTCTAT 1330
Db 841 AT--GAAGTTTCAGGACCTGCAAGAGAGACCTGCGTGGNATGTGAGAGACAGTCTAT 897
QY 1331 CCAATGAGCGTCTCTTGGCCAAACAGCAGGTGTTTCACTACGCTGCTTCCGTTGCTCC 1390
Db 898 CCAATGAGCGTCTCTTGGCCAAACAGCAGGTGTTTCACTACGCTGCTTCCGTTGCTCC 957
QY 1391 TATTGCAACAAACACTCAGTCTAGGAACATATGCACTTTTACATGGAAGAAATCTATTGT 1450
Db 958 TATTGCAACAAACACTCAGTCTAGGAACATATGCACTTTTACATGGAAGAAATCTATTGT 1017
QY 1451 AAGCTCACTTCAATCACTCTTTAAATCTAAGGCAACTATGATGAAGGCTTTGGGCAC 1510
Db 1018 AAGCTCACTTCAATCACTCTTTAAATCTAAGGCAACTATGATGAAGGCTTTGGGCAC 1077

QY 1511 AGACCACACAGGATCTATGGCCAAAGCAAAATGAAAAAGAGATTTTGGAGAGACCA 1570
Db 1078 AGACCACACAGGATCTATGGCCAAAGCAAAATGAAAAAGAGATTTTGGAGAGACCA 1137
QY 1571 GCCAGCTTGCATAATGCAAGGAGACCCCTCACACCCAGGGGTAGAGATCCCCCTATT 1630
Db 1138 GCCAGCTTGCATAATGCAAGGAGACCCCTCACACCCAGGGGTAGAGATCCCCCTATT 1197
QY 1631 GCTAAGGTGGGTGCTCTGGCTGCAAGTATGGAAGCCAAAGCCCTCTCTCAGCAGGAGAG 1690
Db 1198 GCTAAGGTGGGTGCTCTGGCTGCAAGTATGGAAGCCAAAGCCCTCTCTCAGCAGGAGAG 1257
QY 1691 GAAGCAAGCCAGCTGAAACCAAGAGCTGAGGATCGGCTGGCCACCCCACTGAACTT 1750
Db 1258 GAAGCAAGCCAGCTGAAACCAAGAGCTGAGGATCGGCTGGCCACCCCACTGAACTT 1317
QY 1751 GGAAGTTCAGGAAGTGCCTTTCAGAGGAGGGATCAAAATGTCAAAGCCCAATTTGGCTCT 1810
Db 1318 GGAAGTTCAGGAAGTGCCTTTCAGAGGAGGGATCAAAATGTCAAAGCCCAATTTGGCTCT 1377
QY 1811 GAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCTAGATCTGAAGAAGCTA 1870
Db 1378 GAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCTAGATCTGAAGAAGCTA 1437
QY 1871 AGACGATCTTCTTCACTGAAGGAAAGAGCCGCCCATTTCACTGTAGCAGCTTCATTCAA 1930
Db 1438 AGACGATCTTCTTCACTGAAGGAAAGAGCCGCCCATTTCACTGTAGCAGCTTCATTCAA 1497
QY 1931 AGCAGCTCTGTCAAGAGCCCAAAATCTGTGTCGCCACCTATCAGAAAGGCTGGAGCATG 1990
Db 1498 AGCAGCTCTGTCAAGAGCCCAAAATCTGTGTCGCCACCTATCAGAAAGGCTGGAGCATG 1557
QY 1991 TCAGAGCAGAGTGAAGAGTCTGTGGTGGAGAGTTCGAGAAAGGAAACAAGTGGAAAT 2050
Db 1558 TCAGAGCAGAGTGAAGAGTCTGTGGTGGAGAGTTCGAGAAAGGAAACAAGTGGAAAT 1617
QY 2051 GCCAAGGCTTCTTAAGAAAGATGGGAATGTGGGAAAAACAACCTGGCAAAACAAGATCT 2110
Db 1618 GCCAAGGCTTCTTAAGAAAGATGGGAATGTGGGAAAAACAACCTGGCAAAACAAGATCT 1677
QY 2111 AAAGGAGACAGGAGGAGAGAGTAAAGAGGTCTATAGTTTGGAGATGGAGATGAGAA 2170
Db 1678 AAAGGAGACAGGAGGAGAGAGTAAAGAGGTCTATAGTTTGGAGATGGAGATGAGAA 1737
QY 2171 CTTGTAGAAATGGTGCAGACTCCGATGAGATGATACAGCTTCTTCAAACAATCT 2230
Db 1738 CTTGTAGAAATGGTGCAGACTCCGATGAGATGATACAGCTTCTTCAAACAATCT 1797
QY 2231 CCACAAAGAACCCCAAGTCTCTGAAATTTGGTGGAGTTTGTAGCAACACCTTTCTGAAGAA 2290
Db 1798 CCACAAAGAACCCCAAGTCTCTGAAATTTGGTGGAGTTTGTAGCAACACCTTTCTGAAGAA 1857
QY 2291 TTCCTACTCTCAGAAATCAGAAATCCAGGATGTGAACTCTGGAGGGAGAGTGGTCAA 2350
Db 1858 TTCCTACTCTCAGAAATCAGAAATCCAGGATGTGAACTCTGGAGGGAGAGTGGTCAA 1917
QY 2351 GAGCTCTCTGTGGAGAACAGATTAAGAGAAATCGGTATTATGATGAGGATGAGGATGAA 2410
Db 1918 GAGCTCTCTGTGGAGAACAGATTAAGAGAAATCGGTATTATGATGAGGATGAGGATGAA 1977
QY 2411 GAGTGACAAATGCAATGATGCTGGCCTTAAATTTATGTTAGTTAGCGGACCACTGC 2470
Db 1978 GAGTGACAAATGCAATGATGCTGGCCTTAAATTTATGTTAGTTAGCGGACCACTGC 2037
QY 2471 CTTTGTCAAATGTGATGCATTAAGCAGGTATCCAGCAGTGAATGTAAATTTACTTGG 2530
Db 2038 CTTTGTCAAATGTGATGCATTAAGCAGGTATCCAGCAGTGAATGTAAATTTACTTGG 2097
QY 2531 AAGTAACTTTGAAAGAAATTCCTTTTAAATTCAAAAACAAAA 2574
Db 2098 AAGTAACTTTGAAAGAAATTCCTTTTAAATTCAAAAACAAAA 2141

```
RESULT 14
BD160117          2207 bp   DNA       linear   PAT 17-JAN-2003
LOCUS             Primer for synthesizing full-length cDNA and use thereof.
DEFINITION
ACCESSION         BD160117.1 GI:27865875
VERSION           JP 2002191363-A/14960.
KEYWORDS          Homo sapiens (human)
SOURCE            Homo sapiens
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE         1 (bases 1 to 2207)
AUTHORS           Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
                  Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE             Primer for synthesizing full-length cDNA and use thereof
JOURNAL           Patent: JP 2002191363-A 14960 09-JUL-2002;
                  HELIX RESEARCH INSTITUTE
COMMENT           OS Homo sapiens (human)
                  PN JP 2002191363-A/14960
                  PD 09-JUL-2002
                  PF 28-JUL-2000 JP 2000280990
                  PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
                  PI SAITO,
                  PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
                  PI KEIICHI NAGAI,TETSUJI OTSUKI
                  PC
                  C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ 10,
                  C12N5/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ 10,
                  PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
                  Primer for synthesizing full-length cDNA and use thereof FH Key
                  Location/Qualifiers
FEATURES             location
source               1..2207
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
BASE COUNT          692 a 438 c 512 g 565 t
ORIGIN
Query Match          41.0%; Score 1519; DB 6; Length 2207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1569; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1001 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGAGCAAGAGAGAT 1060
DB 638 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGAGCAAGAGAGAT 697
QY 1061 GTGCCCCAGGTCCTGAGGTCTGCATACCCATCAGGAAGGGAAGAAAGATTTCTGCAAT 1120
DB 698 GTGCCCCAGGTCCTGAGGTCTGCATACCCATCAGGAAGGGAAGAAAGATTTCTGCAAT 757
QY 1121 GAGAATAGCTGCGAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTCACTCCAGGTT 1180
DB 758 GAGAATAGCTGCGAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTCACTCCAGGTT 817
QY 1181 AAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 1240
DB 818 AAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 877
QY 1241 AGTCTTTTGAAGTTCTCTCCCAAGCAATGAAGAAAGTTTCAGGCACCTGCAAGAGAG 1300
DB 878 AGTCTTTTGAAGTTCTCTCCCAAGCAATGAAGAAAGTTTCAGGCACCTGCAAGAGAG 937
QY 1301 ACCTGCGTGAATGTCAGAGAGAGTCTATCCCAATGAGCGGCTCTTGGCCCAACAGAG 1360
DB 938 ACCTGCGTGAATGTCAGAGAGAGTCTATCCCAATGAGCGGCTCTTGGCCCAACAGAG 997
QY 1361 GTGTTTCAATCAGCTGCTTCCTGCTCTTCTTGAACAACAACAATCAGTCTAGGAACA 1420
DB 998 GTGTTTCAATCAGCTGCTTCCTGCTCTTCTTGAACAACAACAATCAGTCTAGGAACA 1057
QY 1421 TATGCATCTTTACATGGAAGAAATCTATTGTGAAGCTCACTTCAATCAACTCTTTAAATCT 1480
```

```
DB 1058 TATGCATCTTTACATGGAAGAAATCTATTGTGAAGCTCACTTCAATCAACTCTTTAAATCT 1117
QY 1481 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACACAAGGATCTATATGGCAAGCAAA 1540
DB 1118 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACACAAGGATCTATATGGCAAGCAAA 1177
QY 1541 AATGAAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTGAATGCAAGGGAGACCCCT 1600
DB 1178 AATGAAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTGAATGCAAGGGAGACCCCT 1237
QY 1601 CACAGCCAGGGGTAGAAGATGCCCTATTGCTTAAGGTGGGTGCTCTGCTGCAAGTATG 1660
DB 1238 CACAGCCAGGGGTAGAAGATGCCCTATTGCTTAAGGTGGGTGCTCTGCTGCAAGTATG 1297
QY 1661 GAAGCCAAAGGCTCTCTCTCAGCAGAGAGAAAGAACAGCCAGCTGAAACCAAGAGCTG 1720
DB 1298 GAAGCCAAAGGCTCTCTCTCAGCAGAGAGAAAGAACAGCCAGCTGAAACCAAGAGCTG 1357
QY 1721 AGGATCGCTGGCCACACCCCTCACTTGAAGTTCAGGAAGTTCAGGAAGTTCAGGAAGGG 1780
DB 1358 AGGATCGCTGGCCACACCCCTCACTTGAAGTTCAGGAAGTTCAGGAAGTTCAGGAAGGG 1417
QY 1781 ATCAAAATGTCAAAAGCCCAATGGCTCTGAAGAGCAATCAGCAAGCCAGGATTCCT 1840
DB 1418 ATCAAAATGTCAAAAGCCCAATGGCTCTGAAGAGCAATCAGCAAGCCAGGATTCCT 1477
QY 1841 GAGGATGCGATCTAGATCTGAAGAGCAATGAGATCTCTTCACTGAAGAGCAAGAGCTG 1900
DB 1478 GAGGATGCGATCTAGATCTGAAGAGCAATGAGATCTCTTCACTGAAGAGCAAGAGCTG 1537
QY 1901 CGCCCAATCACTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACCTGTG 1960
DB 1538 CGCCCAATCACTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACCTGTG 1597
QY 1961 TCCCACTATCAGAAAGGCTGGAGCATGTGAGAGAGTGAAGAGTGTGTGGGTGGA 2020
DB 1598 TCCCACTATCAGAAAGGCTGGAGCATGTGAGAGAGTGAAGAGTGTGTGGGTGGA 1657
QY 2021 AGAGTTGAGAAAGGAAACAAGTGGAAATGCCAGGCTCTTAAGAGAAATGGGAATGTG 2080
DB 1658 AGAGTTGAGAAAGGAAACAAGTGGAAATGCCAGGCTCTTAAGAGAAATGGGAATGTG 1717
QY 2081 GGAAAAACAACCTGGCAAAACAAAAGATCTAAAGGAGACAGGAGAGAGAGTAAGGAA 2140
DB 1718 GGAAAAACAACCTGGCAAAACAAAAGATCTAAAGGAGACAGGAGAGAGAGTAAGGAA 1777
QY 2141 GGTCAATAGTTTGGAGATGAGAAATCAGAAATCTTGTAGAAAATGGTGCAGACTCCGATGAA 2200
DB 1778 GGTCAATAGTTTGGAGATGAGAAATCAGAAATCTTGTAGAAAATGGTGCAGACTCCGATGAA 1837
QY 2201 GATGATAACAGCTTCCTCAAAACAACATCTCCACAAGACCCCAAGTCTCTGAATGGTGC 2260
DB 1838 GATGATAACAGCTTCCTCAAAACAACATCTCCACAAGACCCCAAGTCTCTGAATGGTGC 1897
QY 2261 AGTTTGTGAGACAACACCTTTGCTGGAAGATTCCTACTCTCAGAAATCAGAAATCCAGAT 2320
DB 1898 AGTTTGTGAGACAACACCTTTGCTGGAAGATTCCTACTCTCAGAAATCAGAAATCCAGAT 1957
QY 2321 GTGAACTCTGGAGGAGAGAGTGGTCAAGAGAGCTCTGTGGAAGAACAGATAAAGAGA 2380
DB 1958 GTGAACTCTGGAGGAGAGAGTGGTCAAGAGAGCTCTGTGGAAGAACAGATAAAGAGA 2017
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGA CAATTTGCAATGATGCTGGGCTT 2440
DB 2018 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGA CAATTTGCAATGATGCTGGGCTT 2077
QY 2441 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCCTTTGTCAAAATGTGATGACATAGCAG 2500
DB 2078 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCCTTTGTCAAAATGTGATGACATAGCAG 2137
QY 2501 GTATCCCAAGCATGAAATGTAATTTTACTTGAAGTAACTTTGGAAAAGAAATTCCTTCTTAA 2560
```

Db 2138 GTATCCCGAGCATGAATTAATTTACTTTGGAGTAATCTTTGGAAAAGAAATTCCTTCTTAA 2197

Qy 2561 AATCAAAAAC 2570

Db 2198 AATCAAAAAC 2207

RESULT 15

AK023649

LOCUS Homo sapiens cDNA FLJ13587 fis, clone PLACE1009246, weakly similar

DEFINITION co. POLLEN SPECIFIC PROTEIN SP3.

ACCESSION AK023649

VERSION AK023649.1 GI:10435634

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Nagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, F., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T. NEDO human cDNA sequencing project

TITLE NEDO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2207)

AUTHORS Isogai, T. and Otsuki, T.

TITLE Direct Submission

JOURNAL Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- and 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES

source

1..2207

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="PLACE1009246"

/tissue_type="placenta"

/clone_lib="PLACE1"

/note="cloning vector: pME18SFL3"

680..2053

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB14625.1"

/db_xref="GI:10435635"

/translation="MEQENYPPGPEVCIHQEKEISANENSIAVSTPAEDSDRDS QVKEVQVPFAPKLPSPSRASSLSESPPMKKFQAPARETCVCCQTVYPMERLL ANQVHFISCRPCYCNKLSLGTYSLSHGRYICKPHNQLFKSKNGYDEGPHRPHK DLWASKNEEILERPQANARETPSPGVEDAPIAKVGVLAASMEAKASQOEKED KPAETKKLIAPPTPELGGSSGSALEGIKMKPKWPEDEIISKPEVPEDVDLDKLI RRSLSKRSRPFVTAASFQSTSVKSPKTVSPPIKMGMSQESSESVGRVARKQV ENAKSKNGNVGHTQNKESKGETSKRSKSGHSLMENENLVENGADSDSDNFI KQSPQEPKSLNWSFVNTFAEFTTQNKSQDVELWEGEVKELSVBQIKRNYI DEDEDE"

BASE COUNT 692 a 438 c 512 g 565 t

ORIGIN

Query Match 41.0%; Score 1519; DB 9; Length 2207;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1569; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1001 AATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAGAGGAAT 1060

Db 638 AATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAGAGGAAT 697

Qy 1061 GTGCCCCCAGGTCCTGAGGTCTGCATCACCCTCAGGAAGGGGAAAAGATTTCTGCAAT 1120

Db 698 GTGCCCCCAGGTCCTGAGGTCTGCATCACCCTCAGGAAGGGGAAAAGATTTCTGCAAT 757

Qy 1121 GAGAAATAGCTGGCAGTCCGTTCCAGCCCTGCCGAAGATGACTCCCGTGACTCCCAAGTT 1180

Db 758 GAGAAATAGCTGGCAGTCCGTTCCAGCCCTGCCGAAGATGACTCCCGTGACTCCCAAGTT 817

Qy 1181 AAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCCACTAAGTCCAGATTCAGAGGCTCC 1240

Db 818 AAGAGTGAGGTTCAACAGGCTGTCCATCCCAAGCCCACTAAGTCCAGATTCAGAGGCTCC 877

Qy 1241 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCNAGAGAG 1300

Db 878 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCNAGAGAG 937

Qy 1301 ACCTGCGTGGAAATGTGAGAAAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAG 1360

Db 938 ACCTGCGTGGAAATGTGAGAAAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAG 997

Qy 1361 GTGTTTCAATCAGCTGCTTCCGTTGCTCTTATTCGCAACAACAACTCAGTCTAGGAACA 1420

Db 998 GTGTTTCAATCAGCTGCTTCCGTTGCTCTTATTCGCAACAACAACTCAGTCTAGGAACA 1057

Qy 1421 TATGCATCTTTACATGGAAGAATCTATTTAGACCTCACTTCAATCAACTCTTTAAATCT 1480

Db 1058 TATGCATCTTTACATGGAAGAATCTATTTAGACCTCACTTCAATCAACTCTTTAAATCT 1117

Qy 1481 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACACACCAAGGATCTATGGGCAAGCAAA 1540

Db 1118 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACACCAAGGATCTATGGGCAAGCAAA 1177

Qy 1541 AATGAAACAAAGAGATTTTGGAGAGACACGAGCCAGCTTGCATAATGCAAGGAGAGACCCCT 1600

Db 1178 AATGAAACAAAGAGATTTTGGAGAGACACGAGCCAGCTTGCATAATGCAAGGAGAGACCCCT 1237

Qy 1601 CACAGCCCAAGGGTAGAAGATGCCCTTATTTGTAAGTGGGTGTCTCTGCTGCAAGTATG 1660

Db 1238 CACAGCCCAAGGGTAGAAGATGCCCTTATTTGTAAGTGGGTGTCTCTGCTGCAAGTATG 1297

Qy 1661 GAGCCCAAGGCTCTCTCAGCAGGAGAGAGAGCAAGCCAGCTGAAACCAAGAGAGCTG 1720

Db 1298 GAGCCCAAGGCTCTCTCAGCAGGAGAGAGAGCAAGCCAGCTGAAACCAAGAGAGCTG 1357

Qy 1721 AGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTTGAAGGATGCTGAGGAGAGG 1780

Db 1358 AGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTTGAAGGATGCTGAGGAGAGG 1417

Qy 1781 ATCAAAATGTCAAAGCCCAAAATGGCCCTCTGAAAGACGAAATCAGCAAGCCGCAAGTTCT 1840

Db 1418 ATCAAAATGTCAAAGCCCAAAATGGCCCTCTGAAAGACGAAATCAGCAAGCCGCAAGTTCT 1477

Qy 1841 GAGGATGTGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCTACTGAAGGAGAGAGC 1900

Db 1478 GAGGATGTGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCTACTGAAGGAGAGAGC 1537

Qy 1901 CGCCCAATTCAGTGTAGCAGCTTCAATTCAGACCTCTGTCAAGAGCCCAAAACTGTG 1960

Db 1538 CGCCCAATTCAGTGTAGCAGCTTCAATTCAGACCTCTGTGTCAAGAGCCCAAAACTGTG 1597

Qy 1961 TCCCAACCTTATCAGGAAAGGCTGGAGCATGTCAAGCAGAGTGAAGAGTCTGTGGGTGA 2020

Db 1598 TCCCAACCTTATCAGGAAAGGCTGGAGCATGTCAAGCAGAGTGAAGAGTCTGTGGGTGA 1657

Qy 2021 AGAGTTGAG 2080

Db 1658 AGAGTTGAG 1717

Qy 2081 GGAAG 2140

Db 1718 GGAAAAACAACCTGGCAAAACAAGAAATCTAAAGGAGAGACAGGGAAGAAAGTAAGGAA 1777
QY 2141 GGTCTAGTTTGGAGATGGAGATGAGATCTGTAGAAAATGGTGCAGACTCCGATGAA 2200
Db 1778 GGTCTAGTTTGGAGATGGAGATGAGATCTGTAGAAAATGGTGCAGACTCCGATGAA 1837
QY 2201 GATGATAACAGGTTTCCCTCAAAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCG 2260
Db 1838 GATGATAACAGCTTCTCTCAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCG 1897
QY 2261 AGTTTGTAGACAACAACCTTTGCTGAAGAATTCATCTCAGAAATCAGAAATCCCAGGAT 2320
Db 1898 AGTTTGTAGACAACAACCTTTGCTGAAGAATTCATCTCAGAAATCAGAAATCCCAGGAT 1957
QY 2321 GTGGAATCTGGGAGGAGAAAGTGTCAAGAGCTCTCTGTGGAAGAACAGATAAGAGA 2380
Db 1958 GTGGAATCTGGGAGGAGAAAGTGTCAAGAGCTCTCTGTGGAAGAACAGATAAGAGA 2017
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTCGAATGATGCTGGGCCCTT 2440
Db 2018 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTCGAATGATGCTGGGCCCTT 2077
QY 2441 AAATTCATGTTAGTGTAGCGAGCCACTGCCCTTTGTCAAAAATGTGATGACATGAAGCAG 2500
Db 2078 AAATTCATGTTAGTGTAGCGAGCCACTGCCCTTTGTCAAAAATGTGATGACATGAAGCAG 2137
QY 2501 GTATCCACGATGAATGTAATTTACTTGAAGTAACITTTGAAAAGAAATTCCTTCTTAA 2560
Db 2138 GTATCCACGATGAATGTAATTTACTTGAAGTAACITTTGAAAAGAAATTCCTTCTTAA 2197
QY 2561 AATCAAAAAC 2570
Db 2198 AATCAAAAAC 2207

Search completed: January 6, 2004, 20:07:01
Job time : 13116 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:44:24 ; Search time 20 Seconds

(without alignments)
3649.601 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927
Sequence: 1 MESSFPNRQWTSLSLRVTA.....LSVESQIKRNRYDEDEEE 759

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*

1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	246	6.3	5327	T13564	microtubule-associ
2	239	6.1	916	A27864	neurofilament trip
3	232	5.9	734	B42880	nucleolus-cytoplas
4	222	5.7	1313	F96573	hypothetical prote
5	218.5	5.6	200	G84822	probable LIM-domai
6	218	5.6	189	T03400	transcription fact
7	217	5.5	219	S28507	nucleolar phosphop
8	216	5.5	990	I51618	myosin heavy chain
9	215.5	5.5	2139	T18296	hypothetical prote
10	214	5.4	1233	S56271	protein UNC-89 - C
11	213	5.4	6642	T29757	hypothetical prote
12	212.5	5.4	3488	T34418	LIM domain protein
13	212	5.4	211	T47915	probable transcrip
14	212	5.4	226	T02467	hypothetical prote
15	212	5.4	1020	T29108	transcription fact
16	211	5.4	1790	S67593	hypothetical prote
17	211	5.4	1961	A61231	transport protein
18	210	5.3	172	T50694	myosin heavy chain
19	210	5.3	199	T47716	transcription fact
20	207	5.3	852	T06310	transcription fact
21	206.5	5.3	298	T27209	hypothetical prote
22	205.5	5.2	1871	D96796	hypothetical prote
23	205	5.2	849	S00030	probable heat choc
24	204.5	5.2	990	H88733	neurofilament trip
25	204.5	5.2	1091	T34107	protein F32E10.3 [
26	204	5.2	1110	T51116	hypothetical prote
27	203.5	5.2	913	T52485	NF-180 - sea lamp
28	201	5.1	1658	S55101	neurofilament prot
29	201	5.1	2116	A26655	hypothetical prote
					myosin heavy chain

ALIGNMENTS

RESULT 1

T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217689
A:Accession: T13564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A:Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match	6.3%	Score 246;	DB 2;	Length 5327;
Best Local Similarity	22.3%;	Pred. No. 0.00039;		
Matches 192;	Conservative 124;	Mismatches 330;	Indels 214;	Gaps 38;
QY	77	ENPGLGAEHTDSLNRNNGSTEIRHRAHPAEVTSHAAASGAKADQEEQIHPSRLRSPEA	136	
Db	2600	ERPSPAESAKDAA--ESVEKSKDASRPPSVVESTKADSTKGDISPS--PESVLEGPKDD	2655	
QY	137	LVQ---GRYPHIK---DGEDLKDHS-----TESKMKMNCIGESRHE-----VE-----	173	
Db	2656	VEKSKESRPPSVSGASITGDTKDVSRPASVSVSKDEHDKAESRRRSIAKVESVIDEAG	2715	
QY	174	KSEISENTDAGKTEKYNV---PLNRLKMMFEKPEPTQTKI-LRAQSRASGRKISNSY	229	
Db	2716	KSDSKSSQDKDEKSTLASKAESRRSVSVESKDDAKESRPESVIAEGEPVPRESK	2775	
QY	230	S-LDDLEIG-PGQ-LSSTTFDSEKNE--SRNLELPRLSETSIKDRMAKYCAA-----	277	
Db	2776	SPLDSKDTSRPGSVSVESVTAEDEKSEQOQSRRESVAESVKADTKKD--GKSQEA	2833	
QY	278	-----VSKOSSSTNYTNELKASGGITKIHKMOKENVPVPGVCITHQGEKISA	327	
Db	2834	ELLXDDDEKESRRQSITGSHKAMSTWGDSPMDKAD-KSKEPSRPESVAESIKHENTKD	2892	
QY	328	NENSLAVRSTPAEDDSDRSQVKSEV---QQVHPKPLS-----PDSRASSL	370	
Db	2893	EESPLGSR-----DSVAESIKSDITKGEKSPKSVKESRPESVVGSIKDEKAESRRESV	2947	
QY	371	SESSPPKAMKKFOA--PARETC-----VECOKTVYMERLLANQ	408	

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719; PMID:11130712
A:Accession: F96673
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1313 -STO-
A:Cross-references: GB:AE005173; NID:95042434; PIDN:AAD38273.1; GSPDB:GN00141
C:Genetics:
A:Gene: Fl3011.30
A:Map position: 1

Query Match 5.7%; Score 222; DB 2; Length 1313;
Best Local Similarity 20.5%; Pred. No. 0.0011;
Matches 166; Conservative 138; Mismatches 332; Indels 174; Gaps 32;

QY 16 LRYTAKELSLVNNKNSATVEIFSKYKAAEETNME-----KESNTENISQHER--- 65
DB 62 LKKADEQIELLKDKAKAIDDL-KESEKLVVEANEKLEKALAAQKAAEBSFEVEKPRAVE 120
QY 66 --KGTITVLKKWENPGLAESHTDSLNN-----SSTEIRHRAHPHAEVTSAA 113
DB 121 LEQAGLEAVQKK---DVTSKNELEISIQHALDISALLSTTEELQVRVH-ELSMTADAK 175
QY 114 SGAKADQEE-----QIHPSRSLRSPPEALVQ--GRYPHI---KDGBDLKDHSSTESKKNEN 163
DB 176 NKALSHAEAEATKTAETHAEEK---AEILASELGRLLKALLGSKEEKEATEGNEIVSKLKS 230
QY 164 CLESHEHEVKSISENT--DASGKTEKYNVPLNRUKMFMKEGPEPTQTKILRAQSRASG 221
DB 231 EIBLLRGELEKVKVILESSLKEQBLGVLKVDLEAAKM-----AESCTN 274
QY 222 RKXISSENSYSLDDLEIGPGQLSSSTPFSEKNESERNLELPLRSETSKIDRMAYQAAVSKQ 281
DB 275 SSVEEWKWKVHELE-----KEVEENR-----SKSSASESM-----ESWMKQ 311
QY 282 SSSTNTYTNELKASGGGEIKHKMEQENV-----PPGPEVCITHQEGEKISANENSL 332
DB 312 LAELNHVLHETKSDNAAQAEKIELEKTEIAORTDLEEYGRQVCTAKEASKL-----ENLV 368
QY 333 AVSTPAEDDSRDSQVSKVEQVQPHKPLSPPSRASSLESSPPKAMKFKQAPARETC-V 391
DB 369 -----ESIKSELEISQEBKTRALDNEKAATSNIQNLLDQRTSELSIELERCKV 415
QY 392 ECOKTYYPMERL-LANQVPHISC---FRCSYNNKLSLGTVYASLHGRYKCPHPNQLPK 447
DB 416 EEEKSKKDESMETLALQAEASTESSEKATLLVCQBELK-----NCESQVDSLXL 464
QY 448 SKGNVDEGFGHRPHKDLWASKNENBEILERPAAQLANARETPHSPGVGEDAPIAKGVLAAS 507
DB 465 ASKETHEKY-----EKMLEDPARNEIDLSKSTVDSIQNEFNSKA-GWEQKELHMGCVKKS 519
QY 508 MEAKASSQOE-----KEDKPAETKTLRIAWPPPTLGS-SGALEEG 548
DB 520 EENSSSQGEVSRVNLNKKESSEADACARKEEASLKNLKVAGEGVKYLOETLGAKEAS 579
QY 549 IKMSKPKWPPDEISKPEVPEDVDLKKLRR--SSSLKERSRPFPTVAASFOSTSVSKPST 607

Db 580 MKLX-----ESLLKDEEDLKNVTAETISSUREWEGSVLEKIEELS-----KYKES 623

QY 608 VSPPIKGMWSSEOSSESVGGRVARKQVENAKAKKNGKNGVKTWNKSKSGKETGRKSK 667

Db 624 LVDKETKLQSIQTAEELKCGREAAHMQIBELSTANASLVDEATKLSIQVQESSEDLKEKE 683

QY 668 EGH-----SLMEHENIENVGAD--SDSDNSFLUKQSQPOEPKSLNWSFVDNTFAEEFT 720

Db 684 AGYLKKIEELSVANESLADNVTDLQSIQVQESKDKREVAYLKKEELSIVANESLVDKET 743

QY 721 TONKQSDQVELWEGEVVKEL-SVEEQIKRN 749

Db 744 KLQHDQEAELRGREASHLKKIELSKEN 773

RESULT 5

G84822

probable LIM-domain protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 15-Jun-2001

C:Accession: G84822

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.; euss, D.; Nietman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: G84822

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-200 <SOT>

A:Cross-references: GB:A8002093; NID:20886643; PIDN:AAB95275.1; GSPDB:GN00139

C:Genetics:

A:Map position: 2

C:Superfamily: mouse muscle LIM protein MLP; LIM metal-binding repeat homology

Query Match 5.6%; Score 218.5; DB 2; Length 200;

Best Local Similarity 36.5%; Pred. No. 0.00017;

Matches 46; Conservative 23; Mismatches 40; Indels 17; Gaps 4;

QY 387 RETVECOQTVYPMERLLANQQVPHISCFRCSCYCNKLSLGTYSALHGRYCKPHNQLF 446

Db 7 QOKRACEKTVYVELLSADGISVHKACFKCSHCKSLQLSNYSVMGEGVYCRPHFQLF 66

QY 447 KSKGNYDEGFGRPHKDLWASKNEEILERPQAQNA---NARETPHSPGVVEDAIAKV 497

Db 57 KESGSFSKNF-QSPAKPL--TDKTPELNRTPSRLAGM-----PSGTQDKCAICTKTV 118

QY 498 IAKVGV 503

Db 119 IEKTV 124

RESULT 6

T03400

probable transcription factor SF3 - common tobacco

N:Alternate names: LIM-domain protein

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 15-Jun-2001

C:Accession: T03400

R:Shen, W.H.; Gigot, C.

submitted to the EMBL Data Library, February 1997

A:Description: LIM domain-containing protein of Nicotiana tabacum.

A:Reference number: 214931

A:Accession: T03400

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-189 <SHE>

A:Cross-references: EMBL:Y11002

A:Experimental source: Strain Bright Yellow 2

C:Genetics:

A:Gene: SF3

C:Superfamily: mouse muscle LIM protein MLP; LIM metal-binding repeat homology

C:Keywords: DNA binding

F:10-61/Domain: LIM metal-binding repeat homology <LIM>

Query Match 5.6%; Score 218; DB 2; Length 189;

Best Local Similarity 37.2%; Pred. No. 0.00017;

Matches 45; Conservative 24; Mismatches 44; Indels 8; Gaps 3;

QY 387 RETVECOQTVYPMERLLANQQVPHISCFRCSCYCNKLSLGTYSALHGRYCKPHNQLF 446

Db 7 QOKRACEKTVYVELLSADGVNHYKSCFKCSHCKSLQLSNYSVMGEGVYCKPHFQLF 66

QY 447 KSKGNYDEGFGRPHKDLWASKNEEILERPQAQNA---NARETPHSPGVVEDAIAKV 502

Db 67 KESGNFNKNF-QSPAK---SAEKLPELTRSPSKAAGMFSGTQEKCATCGKTAYPLEKVT 122

QY 503 V 503

Db 123 V 123

RESULT 7

S28507

transcription factor SF3 - common sunflower

C:Species: Helianthus annuus (common sunflower)

C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 15-Jun-2001

C:Accession: S28507; S37656

R:Baltz, R.; Domon, C.; Pillay, D.T.N.; Steinmetz, A.

submitted to the EMBL Data Library, February 1992

A:Description: Characterization of a pollen-specific cDNA from sunflower encoding a zinc

A:Reference number: S37656; MUID:93258417; PMID:1302629

A:Accession: S28507

A:Molecule type: DNA

A:Residues: 1-219 <BAL>

A:Cross-references: EMBL:X64392; NID:918918; PID:918919

A:Experimental source: strain HA 401B

R:Baltz, R.; Domon, C.; Pillay, D.T.N.; Steinmetz, A.

Plant J. 2, 713-721, 1992

A:Title: Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger

A:Reference number: S37656; MUID:93258417; PMID:1302629

A:Accession: S37656

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-168, 'E', 170-219 <BA2>

A:Cross-references: EMBL:X64392

A:Note: 169-Lys was also found

C:Genetics:

A:Gene: SF3

A:Introns: 46/3; 79/1; 92/3; 122/3

C:Superfamily: mouse muscle LIM protein MLP; LIM metal-binding repeat homology

C:Keywords: DNA binding; zinc finger

F:11-62/Domain: LIM metal-binding repeat homology <LIM1>

F:110-161/Domain: LIM metal-binding repeat homology <LIM2>

Query Match 5.5%; Score 217; DB 2; Length 219;

Best Local Similarity 35.8%; Pred. No. 0.00023;

Matches 48; Conservative 23; Mismatches 45; Indels 19; Gaps 4;

QY 379 MKKFOAPARETCVECOQTVYPMERLLANQQVPHISCFRCSCYCNKLSLGTYSALHGRYCK 438

Db 1 MKSGTGTQK-CTVCEKTVYLVKLVANQVRVYHKACFCHCHCNSTLKLNSFNPSFGVYVC 59

QY 439 KPHFNQLFPSKNGYDEGGRPHKDLWASKNEEILERPQAQNA---NARETPHSPGVVEDAIAKV 496

Db 60 RHFFDQLFKRTGSLKSPDGT-----KPKPERTFSQETQSAN-RLSSPFEGTRDKCN 111

QY 497 -----PIAKGV 503

Db 112 ACAKIVYPIERVKV 125

RESULT 8

IS1618

nucleolar phosphoprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Sep-1999
C:Accession: F51618; S57757
R:Cafrns, C.; McStay, B.
J. Cell Sci. 108, 3339-3347, 1995
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNoppl80
A:Reference number: F51618; MUID:96019267; PMID:7593294
A:Accession: F51618; S57757
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-990 <EHL>
A:Cross-references: EMBL:X80927; NID:g895920; PIDN:CAA61368.1; PID:g895921
C:Genetics:
A:Gene: xNoppl80
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 5.5%; Score 216; DB 2; Length 990;
Best Local Similarity 19.6%; Pred.No. 0.0016;
Matches 152; Conservative 115; Mismatches 272; Indels 238; Gaps 29;

QY 20 AKELSLV-----NNKSSAIVEIFSKYQAAETNMKKSENLENLSQHFRKGLTLVLYKK 74
DB 28 AKFSKATGVKNDSCPTSLDIFSDWVKPD-----AKKRPAN-----GLPKK 73
QY 75 KWENPGLGASHSTDLRNSSTETIRHRADHPHAEVYTHAASGAKADQEEQIHPRSLRSP 134
DB 74 K-----SAKESSSSEDSSSE-----EDEPPAKBAQPAQGGK-----PVVKAQPK 114
QY 135 EALVQGRYPHIKGGEDLKHSTESKK-----MENCLGESRHEVEKSEISENTDAS 184
DB 115 KAKSSS-----EDSSDESDEEETKPPAKRPAQTPKVAAVAKTPTQKAKXSSSESSESSE 169
QY 185 GKIEKYNVPLNRLKMMKEKEGPTOTKILRAQSRASORKISENSYSLDDLEIFGQLSSS 244
DB 170 DEAKKKQPVIKV-----PPQAAVYKAGLASNNKG--TAPSSSESDSDSPAKKTA 219
QY 245 TFDSEKNESRNLLEPLRLSETSIDRMAKYQAAVSKOSSSTNYTNELKAGGGIKIHKME 304
DB 220 TKTP-----PTKPATAAKPOAKK---TAGKXSSSRESDS--SDEQKTAASK 263
QY 305 QKEN-----VPPGPVCTIHOGEKISANENSIAVSTPAEDDGRDSQVKS-EVQQPVH-- 357
DB 264 PKPDVSAVPPPTSV-----SKKXTLSQPGTKAKPESSDSSSDSDEEQPAKKA 312
QY 358 -----PKPLSPDSRASSLSSESPPKAMKKFQAPARETCVCEQKTVTPMERLLAN 406
DB 313 KIVPAKAAASAPKALAKAETSTDSESDSSSEDEK----- 347
QY 407 QQVPHISCFRCVCYNKLSLGTVASLHGRIYCKPHFNQLFKSGNYDEGFGRPHKDLWA 466
DB 348 -----KSSVKLVGRKAA-----PKK----- 361
QY 467 SKHENEELERPAQLANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKP-AET 525
DB 362 -----APA-APDAKSTPVAACKSAPAKA---SSSSSDSSSNEETTKPAKT 407
QY 526 KKLRIAWPPTELGSSGSALEEGIKMSKPKWPEDEISK-----PEVPEDVDLKLKL 578
DB 408 TPAKSAATPTSKTPTNGKATPTSKTAPKPTPTSTAKKSSSSSDSSSDSDETTTKPA 467
QY 579 RRSSLLKERSRPTTVAASFQSTSVKSPKTVSPP-----IRKGMWSQSEBSVGRVA 631
DB 468 AKTTPAKSAATPTSKTPTNSKATPTSKTTPAKPGTPTKTSAAKKDSSSSSDSSSDEKKT 527
QY 632 ERQOVENAKASKXNGNVGKTTTWQKESKGETGK---RSKEGHSLEVENENLVENGADSD 688
DB 528 PAKRAAKTTTPAKP---AAKTTPAKPAKTTTPAKPAKSTFGKQVPTKXSSSSSDSSSS 584
QY 689 DDNSFLKQSQPQP-----KSLNWSSFVDNTFAEFTTQNOK 725
DB 585 ED-----EKKGSAKPAVKTTTGKATSKPVVAKFPVAKKAKSSSDSSSDEETTKTK 617

QY 634 KOVENAKSKNGVNGKTTWQ-----NKESKGETGKRSKEGHSLEMENLV-ENGAD 685
DB 1706 KKIQ-AELDEVKFNLEDVTNQREKLVAKNGENDAEIDSLKEEKALDETEKITDDNKL 1764
QY 686 SDEDDNSELKQSQFQPSLWNS---SFVNTFAEBEFTTQNKQSQDVLMW--EGEVVKEL 740
DB 1765 SEEDSLDRKYNALLDSKSDVSMKRFQDELKVTQKQALETEKHAETKRLKGRLEKE- 1823
QY 741 SVEEQIK 747
DB 1824 AAEVOVR 1830
RESULT 10
S56271
hypothetical protein YFR016c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 19-Apr-2002
C:Accession: S56271
R:Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasano
submitted to the EMBL Data Library, May 1995
A:Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces ce
A:Reference number: S56186
A:Accession: S56271
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1233 <MUR>
A:Cross-references: EMBL:D50617; NID:9836685; PIDN:BA09255.1; PID:d1009896; PID:g836771
C:Genetics:
A:Cross-references: SGD:S0001912
A:Map position: 6R
Query Match 5.4%; Score 214; DB 2; Length 1233;
Best Local Similarity 20.5%; Pred. No. 0.0027;
Matches 183; Conservative 125; Mismatches 308; Indels 278; Gaps 46;
QY 43 KAABETNMKRSN-TENLSQFRKG-----TLTVLKKWENFQGAESHTDS 89
DB 334 KDVSSESLTKNGFNKFNKESKHLXAGEKQOTESDRDGISPSVLAKNQKETEIGKEDHVE 393
QY 90 LRNASTE-----IRHADHPAETVTHAASGAKAD-----QEEQIHPKSR 130
DB 394 QKXEDKCRKELSVNHENN-----NSHNFAAGSDSIIPPERETYDDTGMPTKRI 447
QY 131 RSPPEALVQR-----YPHIKQEDLKHSTESK-KVENCIGES-----RHEV---EKS 175
DB 448 SDNEKNLQHGNDISVEVEKEEBEENSTFSKVENVTGCEQFAVRNVEYSGTEEE 507
QY 176 EISENTDASGKIEKYNVPLNLKWMFKGPTQTKILRAQSRASGRKISSENSYSLDLE 235
DB 508 STSKGEEIMGDEKQS-----ZAGE--KSIIEIEG-SANSAKISKONLVLEDEA 554
QY 236 IGPQLSSST-----FD-----SEKNESRRNLPL-----RLSE--T 265
DB 555 EAPTOENKPTVEGIDIPADPRDDVEIYEAVERNIIPEDEVAKQEQGEQVQLDEPVK 614
QY 266 SIKRMAYQAVS-----KQSSNTYNEL-----KASGGEIKHKMQKE 307
DB 615 AWKDDKIAMRGAESISEDMKKQGTAELSNEKAKKEYDETARESAGVEV-----EKS 668
QY 308 NVPPQPEYVC--ITHQEGEKISANENSLAVRSTPAEDDSR--DSOVKSEV----- 352
DB 669 KTPSPKVKVCTSGRPDLQINERDPV---LKEDVRVDEVDVKPEIATTIENSEED 724
QY 353 -----QOPVHPKPLSPDRASSLSSESPPKAMKFFQAPARETCVQCQTVYPM 401
DB 725 PKSORVOISTEQAETTKQDMGDVGSTTSFKEEKPKRFEITQEGDKITGD----- 775
QY 402 RLLANQVFIHSCFRCSCVNNKLS-LGTYASLHRIYKPKHFNQFVSKNGNYDEFGHRP 460
DB 776 ---TNHE--HGEATEAAGENSADSVGT-----AEKVIEPSSSEVKD-----T 814
QY 461 HKDLWASKNENEIILERAQLANARETHPSGV-----EDAPI----- 498

DB 815 BEDAEVENSEKTEF:KVKAELNL-DAPKEAEVTAELNKENEDVEDVTEDEAVENSEKT 873
QY 499 -----AKVGVLAAAMEAKASSQOEKEDPKAETKKLRIAWPPPTTELUGSSGSALEEGIKWS 552
DB 874 EPIKVAELGNLDAPKEAEVTAELNKENEDVEV-----AATSKEDIETKCS 919
QY 553 KPKWPP-----EDEISKPE---VPEDVDLDLKKLRSSSLK-----ERSRPTVAAS 596
DB 920 EPAETPIEDGTCTEAEVSKKDAEAVTKEDENWENSKI--AEALKDVTGQOEIDINISDE 977
QY 597 POSTSVKSPKTVSPPIRKQWSMSQSESESVGGRVRAERKOVENAKASKNGNVGKTTWQ- 655
DB 978 FO-----RTVELP-----ELEKQDIKDKGDEKELEVEETEKETSLPDLVVEENI 1022
QY 656 KESKGETGKRSKEGHSLEM-ENENLVENCADSDDEDDNSFLKQSQSQEPKSLN-----WSSF 710
DB 1023 TEKEIKOKEEBEVSQDFNETESISKAPNND--NGFEDQSTRENPKKASADDIFKDI 1080
QY 711 VNTTFAEBEFTTQNKQSQDVLMWEGVWKLVSVE-----EOIKKNRYVDEDED 757
DB 1081 LDET--NEFLEQLKIVDUSEL--NALLQSLADKADSTTQTTQESKKN--DKQD 1128
RESULT 11
T29757
protein UNC-89 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C:Accession: T29757
R:Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid C09D1.
A:Reference number: Z20679
A:Accession: T29757
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6642 <DUZ>
A:Cross-references: EMBL:AF003131; PIDN:AA54132.1; GSPDB:GN00019; CESP:unc-89
A:Experimental source: strain Bristol N2; clone C09D1
C:Genetics:
A:Gene: CESP:unc-89
A:Map position: 1
A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1;
3/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1
Query Match 5.4%; Score 213; DB 2; Length 6642;
Best Local Similarity 21.3%; Pred. No. 0.024;
Matches 178; Conservative 97; Mismatches 325; Indels 234; Gaps 36;
QY 21 KELSUNKNKSSAIVEIFSKYQAAAEETNMKKRNTENLSQHFRKGLTVLKKWENPG 80
DB 1085 KQSDRVEIREFDGSIKISIKNIKIEDAGEIRAVATNSEGSD--TKAKLTVQKKPF--- 1138
QY 81 LGAESHTDSLNSSTEIRHRAHPPAEVTSHAASAKADQOEQIHPRSLRSPPALVQ 140
DB 1139 ---APEFLRPSVLTVEKSE---AVSAHAFGIPLPTYEWSVNGR-KVRDQEGART 1190
QY 141 RYPHIKDGEDL-----KDHSTESKKNENCLG-----ESRHS---VEKSE 176
DB 1191 RRESTVDGASILITDATTYSEVNHLTISVVAENTLGAEETGAQLTIEPKESVVERKOD 1250
QY 177 ISGNTDASGKIEKYNVPLNRLKWMFKGEP-----TQTKIL----- 212
DB 1251 LS-----SSEVQK-----EIAQVKEASPEATTTITWETSLSSTKTMTTSTTEVTVG 1299
QY 213 -----RAQSRAS-----GRKISENSYSLDDEIPIGQQLSSSTTFSEKESRR--- 255
DB 1300 GVTVEKSESESATTVIGGSGGVTEGISVSKIEV-VSKTDSQTDVREGTPKRRVSFA 1358
QY 256 NLELPELSTSKDRMAYQAAVSKQSSNTVNTLKSAS-----GGEIKHWEKENV 310
DB 1359 EELPK--EVIDSDRKKKSPDKKESPKTEBKPASPTKKTGEEVK-----SPKESKP 1412

QY 311 POPEVCITHOGEKISANENSIAVRSTPAEDDSDRSQVSEVQVPHPKPLSPDSR-ASS 369
DB 1413 ASP-----TKKESPAEE-----VKSPTKKKESPSPTKKE-----KSPSPTKTKTDE 1457
QY 370 LSESSPKA-----MKKFOAPARETCVE--CQKTVYPMERLLANQOVF 410
DB 1458 VKEKSPKSPKTKESPEKPEDVKSPVKKESPDATNIVEVSSETTIEKTTMTTEN-- 1515
QY 411 HSCFRCSYNNKLSLGTYSALHGRYCKPHFNQLFKSGNYDEGFGHRPHKDLWASQNE 470
DB 1516 -----THESBESRTSVKKE-----KTPKVDKPKGPTKKDKSPEKSI 1553
QY 471 NREIL-----ERPAQLANARETHSPG-----VEDAPIAKVG 502
DB 1554 TBEIKSPVKKESPEKVEKPAKSPKTKKESPKASPTKKSENVKSPKTKKESKSP--EKS 1611
QY 503 VLAASNEAKASSQOEKEDKP-AETKKLRITAMP-----PTELGSSGSALDE-- 547
DB 1612 VVEELKSPKESPEKADDPKSPKTKK-----SPEKSATEDVKSPKTKKESPEKVEKETS 1668
QY 548 -GIKMSKPKWPPDEIRSKPEVEDVDLKLRLRRSSSLKERSRPTVAASFQSTSVKSPK 606
DB 1669 PTKKSSPTKTDDEKSPKTKKESKSPQTVKPKASPTKKEKSPKSVV-----BEVKSFK 1723
QY 607 TVSP-----PIRKGWSMEQSEESVGGRAERKQVENAKSKNGNVGKTTWQNK 657
DB 1724 EKSPKAEKPKSPKTKKESPEKSAEYKSPKTKKESPEKSAEYKSPKTKKESPEKSPVKM 1783
QY 658 SKGETCKRSGEHSLE-----MENENLVENGADSDDEDNSFLKQOSPOEP 702
DB 1784 ADDEVKSPKTKKESPEKVEKPAKSPKTKKETPEKSA-AEELKSPKTKKESKSPSP 1836
RESULT 12
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F12F3.
A:Reference number: Z21521
A:Accession: T34418
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3488 <FUL>
A:Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
A:Experimental source: strain Bristol N2; clone F12F3
C:Genetics:
A:Gene: CESP.F12F3.3
A:Map position: 5
A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1
Query Match 5.4%; Score 212.5; DB 2; Length 3488;
Best Local Similarity 19.5%; Pred. No. 0.012;
Matches 155; Conservative 129; Mismatches 328; Indels 183; Gaps 27;
QY 39 SKYQKAAETNMEKRSNTENISQHFRTKGTTLVLKKNWENPGLGAEHSHTSLRNSTEIR 98
DB 716 SKLEKAADTT---KQIETETVDD--KSKKKVKKKTE-----KSDSPTS 755
QY 99 HRADHP--REVTSASHGA-----KADQEQIHRP-----SRLRSPPEA 136
DB 756 QKSETPVVVEPTKPAESBAQKTAENVKAKQKQEVDDNLKREAEVAAKKIADSKLIEAEA 815
QY 137 LVQGRYPHDKGDDLDKDHSTESKVMENCLGSRHEVEKSEIGENTD---ASGKIEKYNV 192
DB 816 NIK-KTAEVEAAKKQE-KDEQLKLETEVTVKSKAAEKLKLEKQAKKAADAVKQK 873
QY 193 PLNRLXWMEKFBPTQTKILRAQSRASGRKISSENSYSLDDLEIGPGLSSSTFDSEKNE 252

DB 874 ELNEKNKLEAAKSAADKLEBESAAKSKVSESVKFFGEKTKTAGEKTVQVSEPT- 932
QY 253 SRNLELPRLSETSIKDRMAKYQAAVSKOSSSTNYTNELKASGEIKIHMEQKENVPPG 312
DB 933 SKTIDTKVGATPEADETPKKIIKKKTEKSSSISQKSATDSE-KVSQKQEODE-PTK 990
QY 313 PEVCITH-----OGEKISANENSIAVRSTPAEDDSDRSQVSEVQVPHPKPLSPDSRA 367
DB 991 PAVSETQMTVEADKSKKQKQKEDKLDLDAELIAKTKQEADEKSKLDAEKIKKVSDEDDAA 1050
QY 368 SSLSE-----SSPPKAMKFOAPARETC-VECOKTVYPMERLLANQOVFH 411
DB 1051 RKKEKLDKLESEIATKKASADKLEBQAQAKKAAAEVAAK----- 1094
QY 412 ISFCRCYNNKLSLGTYSALHGRYCKPHFNQLFKSGNYDEGFGHRPHKDLWASQNE 471
DB 1095 ---KQEKDEQLKLDTEAA-----SKKAAAEKLE- 1120
QY 472 EEILPERPAQLANA-----BETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDK- 521
DB 1121 ---LEKQAOIKKAAGADAVKKQKLEDEKNLEANKSAAGLKIEBESAAKSKQTVBEOA 1177
QY 522 --PAETKKLRITAMPPTTELGGSSGSALEEGIKMGKPKWPPDEISKPEVPEDVDLKLKL 579
DB 1178 KLDATK-----AKTAEKQTKLEKDEKSTKES---ESKETVDEKPKKKVKKKTEK 1225
QY 580 RSSSLKERSRPTVAASFQSTSVKSPKTVSPPIRK-----GHSMSQSE 623
DB 1226 SDSSISQKSETKTVBSAGPSESETQKADAAARKQKETDEKQKLEAEITAKKSADEKSK 1285
QY 624 ESVGGRVAERKQVENAKSKNGNVGKTTWQNKESKQETGKRSKE--GHSLEMENENLVE 681
DB 1286 LEABSKLKAAAEVAAKKQE-----KDEQLKLDTEAAKSKAAAEKLKLEKQSHIK 1336
QY 682 NGADSDDEDNSFLKQOSPOPKSLNWSFVNDTFAEEFTTQNKOSQDVLEWEGEWKELS 741
DB 1337 KAAEVD---AVKQKLEBKQRLSEPAATKADAEXKLEBKQKKAETALIEIQE-- 1390
QY 742 VEEQIKERNRYDEDE 756
DB 1391 -QEKLAQEQSRLEDE 1404
RESULT 13
T47915
LIM domain protein - Arabidopsis thaliana
N:Alternate names: protein T20K12.130
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 15-Jun-2001
C:Accession: T47915
R:De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quer
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z24480
A:Accession: T47915
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <DEH>
A:Cross-references: EMBL:AL1337898
A:Experimental source: cultivar Columbia; BAC clone T20K12
C:Genetics:
A:Map position: 3
A:Introns: 46/3; 80/1; 89/3
A:Note: T20K12.130
C:Superfamily: mouse muscle LIM protein MLP; LIM metal-binding repeat homology
Query Match 5.4%; Score 212; DB 2; Length 211;
Best Local Similarity 26.6%; Pred. No. 0.0039;
Matches 47; Conservative 29; Mismatches 47; Indels 54; Gaps 2;
QY 371 SESSPPKAMKFOAPARETCVECOKTVYPMERLLANQOVPHISFCRCYNNKLSLGTYA 430
DB 86 SDAAPNRLSFFSGTQDKCAACKTKTVVPLEKMTMEGESYHKTFCRCAHSGCLTHSSYA 145

```

QY 431 SLHGRIYCKPHNQIFKSGKGYDEGFGHRPHKDLWASKNENEIILERPAQLANARETPHS 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 146 ALDGVLYCKVHFSQLFLEXGNYNH-----ANFTEETSDAAAE 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 491 PGVEDAPIAKVGVLAAASMEAKASSOEKDEPAETKCLRIAWPPPTLGGSSGSALEE 547
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 -----VLQAANHRRSTAEEDKTEPKED-----ANFTEETSDAAAE 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
T02467
Probable transcription factor SP3 F4118.22 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 15-Jun-2001
C:Accession: T02467; A84895
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC F4118 genomic sequence.
A:Reference number: Z14674
A:Accession: T02467
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-226 <ROU>
A:Cross-references: EMBL:AC004665; NID:g3386593; PID:g3386614
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, J.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84895
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-226 <STO>
A:Cross-references: GB:AE002093; NID:g3386614; PIDN:AAC28544.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g45800; F4118.22
A:Map position: 2
A:Supraons: 45/3; 88/3
C:Superfamily: mouse muscle LIM protein MLP; LIM metal-binding repeat homology
F:10-61/Domain: LIM metal-binding repeat homology <LIM>

Query Match 5.4%; Score 212; DB 2; Length 226;
Best Local Similarity 39.8%; Pred. No. 0.00042;
Matches 39; Conservative 16; Mismatches 33; Indels 10; Gaps 2;

QY 388 ETCVECKTVYPMERLLANQOVPHISCFRCISYNNKLSLGTYSALHGRIYCKPHNQIFK 447
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 DKKACDKTVYVMDLLTLEGNTYHKSCFRCTHCKGLTVISNYSMMGDLVLYCKPHFEQLFK 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 448 SKGNYDEGFGHRPHKDLWASKNE--NEEILERPAQLAN 483
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 ESGNYSKNF-----QAKTEKPDNDLHTRTPSKLSS 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
T29108
hypothetical protein - Trichomonas vaginalis (fragment)
C:Species: Trichomonas vaginalis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T29108
R:Meysick, K.C.; Garber, G.E.; Dimock, K.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z20572
A:Accession: T29108
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: mRNA
A:Residues: 1-1020 <MEV>
A:Cross-references: EMBL:U44915; NID:gl177869; PID:gl177870; PIDN:AAA86838.1
A:Experimental source: clinical isolate

Query Match 5.4%; Score 212; DB 2; Length 1020;

```

Best Local Similarity 20.1%; Pred. No. 0.0027;	
Matches 163;	Conservative 122; Mismatches 306; Indels 218; Gaps 33;
QY	19 T A X E L S I L V N K N K S S A I V E I F S Y Q Y A A E E T M W E K R S T N E N L S Q H F R K G T I T V L K K K W - 77
DBb	121 T A T I S I L S K V N L K P A D N L L A K - - - S Q E N E N K A E S N T D S - - - K P K L T A M N Q F S L 170
QY	78 - - - - - N P G L A E S H T D S L U R N S T E - - - I R H R A D H P A E - - - - - V T S H A A S G A K 117
DBb	171 K P G T P N S E A K P A I G N G L G N K L S S S P L S N I S E A A Q K P A T G G L S K L P L S S K P L S N I Q 230
QY	118 A D O E E Q I H P R S L R - - - S P E A L V O G R Y P H I K Q E D E L - - - K D H S T E S K - K W E N C L G E S R H 170
DBb	231 A A S E E K S Q S T G L K G M K L P S N Q A S E E K P - K L G N L P K P S N S T E E K P K L G N L G K S P S 288
QY	171 E V S K S E I S - - - - - E N T D A S G I E K Y N Y P L N R L K M F E K G E P T O T K I L R A Q S R A S G R K I 224
DBb	289 T E E K P K L Q G I K I G O T P S N S T E E K P K L G N L P K S P S T E E K P K L S N L G K S P S T E E K P K L 348
QY	225 S E N S Y S L D D L E I T G P G O L S S T T P D S E K N S R R N L E P L R L S E T S I K D R M A K Y Q A A V S Q O S S 284
DBb	349 S - - - - - L N L G K S P N O S T - - - - - E E K P K L G N L P K S P S Q T S E E K P K L S N L G K S P S N 396
QY	285 T N Y T N E L K A S G G E I K I H K V E Q E N V P P G E V C I - - - - - T H O E G E K I S A N E N S L A V R S T P A 339
DBb	397 Q S T E E K P K L Q G I K I G O T Q S N O K S E E K P K L S N L G K S S T E E K P K L S N - - - - - L G K S P S 451
QY	340 E D D S R D S Q V K S E V Q Q V H P - - - - - - - - - - - K P L S P D S R A S S L - - - - - 370
DBb	452 - N Q S T E E K P K L G F N L P K A P S N O T E E K P K L G T G G I S L N L G N K P Q S E E K P K L S L G G I K L A Q S 510
QY	371 - - - - - S S S P P K A M K F Q A P A R T C V E C O K Y T V P M E R L L A N Q O V F H I S C F R C S Y N N K L S L 426
DBb	511 P S N S E E K P K L S I N L P K S P N S T E E K P K L G F N L P K A P S N Q - - - - - - - - - - - T E E K P K L 558
QY	427 G T V A S L H G R I Y C K P H F N Q L F K S K G N Y D E G F G H R H K D - - - - - - - L W A S K N V E E I 474
DBb	559 G T - - - - - - - - - - - G G I S L N L G N K P Q S E E K P K L S L G G I K L A Q S P S N S N E - - - 595
QY	475 L E R P A Q L A N A R E T P H S P G V E D A P I A K V G Y L A A S M E A K A S S Q O E K E D X P A E T K L R I A M P P 534
DBb	596 - E X E K L S N L P K S P S N Q S T E E K P K L Q G I K L N L G N K P Q T Q T O T E E K P - - - K L Q L G - - - 647
QY	535 P T E L G S S G S - A L B E - - - - - G I M K S K P W P E D E I S K P E V P E D V D L D L K L R S S - - S 583
DBb	648 G I K L G O S P S N S T E E K P K L Q G I K L N L G - - - - - S K P Q T E E K P K L Q G I K L G T G G I S 699
QY	584 L K E R S R P T V A A S F Q S T S Y K S P K T V S P P I R K G W S E Q S E S V G R V - - - A E R K O V E N A K 640
DBb	700 L N L G N K P - - - - - - - Q S E K P K L Q G I K L G N S O P N O P L E K P K S G I N L N L G K S Q P S S E E 750
QY	641 A S K N G N V G T T W Q N K S K E T G K R S K E G H S L E M E N E N L V E N G A D S D E D D N S F L K Q O S P Q 700
DBb	751 K P K L G L N L G S P S N S T E E K P K L G - - - T G G I S I N L G N - - - - - - - - - - - K P Q T E E 769
QY	701 E P K - S L W N S S F V D N T F A E F T T Q N O K S Q D 728
DBb	790 K P K L S I N L P K S P S N - - - - - - - Q N O S T E E 810

Search completed: January 6, 2004, 09:47:53
Job time : 24 secs

Query Match 5.4%; Score 212; DB 2; Length 1020;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:46:24 ; Search time 42 Seconds
(without alignments)
4663.379 Million cell updates/sec

Title: US-09-890-549-4
Perfect score: 3927
Sequence: 1 MESSPFRQWTSLSLRVTA.....LSVEEQIKNRYDEDEEE 759

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:**

- 1: sp_archaea:
- 2: sp_bacteria:
- 3: sp_fungi:
- 4: sp_human:
- 5: sp_invertebrate:
- 6: sp_mammal:
- 7: sp_mmc:
- 8: sp_organellie:
- 9: sp_phage:
- 10: sp_plant:
- 11: sp_rodent:
- 12: sp_virus:
- 13: sp_vertebrate:
- 14: sp_unclassified:
- 15: sp_rvirus:
- 16: sp_bacteriap:
- 17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2970	75.6	753	11 Q8K2H0	Q8k2h0 mus musculu
2	2955	75.2	753	11 Q8C7S2	Q8c7s2 mus musculu
3	2395	61.0	593	11 Q8CD09	Q8cd09 mus musculu
4	2390	60.9	593	11 Q8BT15	Q8bt15 mus musculu
5	2384	60.7	593	11 Q8C3R7	Q8c3r7 mus musculu
6	1128	28.7	629	13 Q9DEY8	Q9dey8 brachydanio
7	641	16.3	951	6 Q8HXJ9	Q8hxj9 macaca fasc
8	529	13.5	519	4 Q8N7Z0	Q8n7z0 homo sapien
9	363	9.2	128	11 Q8BG05	Q8bg05 mus musculu
10	357	9.1	127	4 Q9BT23	Q9bt23 homo sapien
11	357	9.1	127	4 Q96S91	Q96s91 homo sapien
12	258	6.6	2081	10 Q9LH98	Q9lh98 arabidopsis
13	251	6.4	5412	5 Q9W596	Q9w596 drosophila
14	246	6.3	5327	5 Q76E91	Q76e91 drosophila
15	241.5	6.1	7210	5 Q9V7G8	Q9v7g8 drosophila
16	241.5	6.1	9270	5 Q8MLD9	Q8mld9 drosophila

17	231	5.9	3111	5 Q9VH10	Q9vh10 drosophila
18	223.5	5.7	2083	5 Q9N435	Q9n435 caenorhabdi
19	223.5	5.7	10578	5 Q8ISF5	Q8isf5 caenorhabdi
20	223.5	5.7	18519	5 Q8ISF6	Q8isf6 caenorhabdi
21	223.5	5.7	18534	5 Q8ISF7	Q8isf7 caenorhabdi
22	223	5.7	2465	5 Q8IQ71	Q8iq71 drosophila
23	222	5.7	1313	10 Q9XI96	Q9xip6 arabidopsis
24	222	5.7	1420	5 Q8INM3	Q8inm3 drosophila
25	222	5.7	1430	5 Q9VGZ9	Q9vgz9 drosophila
26	221	5.6	4723	5 Q8MUJ8	Q8muj8 drosophila
27	220	5.6	811	4 Q9P2I3	Q9p2i3 homo sapien
28	218.5	5.6	200	10 Q04193	Q04193 arabidopsis
29	218	5.6	189	10 P93356	P93356 nicotiana t
30	217	5.5	214	10 Q9SP34	Q9sp34 helianthus
31	216.5	5.5	1135	6 Q8HZW3	Q8hzw3 canis famil
32	216	5.5	990	13 Q91803	Q91803 xenopus lae
33	216	5.5	1253	5 Q8SXM6	Q8sxm6 drosophila
34	216	5.5	1298	5 Q9VEU8	Q9veu8 drosophila
35	215.5	5.5	873	10 Q8GUP3	Q8gup3 arabidopsis
36	215.5	5.5	2139	5 Q07569	Q07569 entamoeba h
37	215	5.5	188	10 Q9SP33	Q9sp33 helianthus
38	215	5.5	985	5 Q9U0S5	Q9u0s5 mytilus gal
39	215	5.5	1705	5 Q9U0S7	Q9u0s7 mytilus gal
40	214.5	5.5	2910	10 Q9FND5	Q9fnds arabidopsis
41	214	5.4	188	10 Q9XRG9	Q9xhg9 helianthus
42	214	5.4	1226	4 Q9NZB6	Q9nzb6 homo sapien
43	213	5.4	878	13 Q8AVW4	Q8avw4 xenopus lae
44	213	5.4	6632	5 Q01761	Q01761 caenorhabdi
45	213	5.4	6632	5 Q17362	Q17362 caenorhabdi

ALIGNMENTS

RESULT 1
Q8K2H0
ID Q8K2H0 PRELIMINARY; PRT; 753 AA.
AC Q8K2H0;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC IONS.
CC EMBL; BC031490; AAH31490.1; -.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00478; LIM DOMAIN 1; 1.
DR PROSITE; PS00023; LIM DOMAIN 2; 1.
KW LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 753 AA; 84059 MW; 79F98B47C100CF22 CRC64;

Query Match 75.6%; Score 2970; DB 11; Length 753;
Best Local Similarity 75.9%; Pred. No. 5.1e-168;
Matches 577; Conservative 60; Mismatches 115; Indels 8; Gaps 3;

QY	1	MESSPFRQWTSLSLRVTAKELSLWNKNSAIVEIFSKYKAAAEETNMEKRGNTENL 60
DB	1	MESTFPNRQWTSLSLRVTAKELSLWNKNSAIVEIFSKYKAAAEETNMEKRGNTENL 60
QY	61	SOHFRTGLTTLVKKWENPGLGAESHTDLSNLSPEIRHRADHPPEVTTSHAASGAKADQ 120
DB	61	POHFRGTLVLKKWENPVGAEPHTDLSLPNSSEGHTADYPPAEVTDKFAFGVRAD 120

```

QY 121 EQIHPRSRRLSPPEALVQGRYPHIKQGEDLKORHSTESKMKMENCIGSRHEVEKSEISEN 180
DB 121 EEHTQPKRPFGRREAVIQSRYPSENSHDPKFAQATESQKMKENCIGDGRHEAEKPEISEN 180
QY 191 TDASGKIKYVNPVNLRLKMWFEKEPOTQTKILRAQSRASGRKISENSYSLDDLEIGPGQ 240
DB 181 TETSGKIEKYVNPVNLRLKMWFEKEGHNTQKSLWTSQSRNAGRRRLSENNSLDDWEIGAGH 240
QY 241 LSSSTFDESKNESRRNLPLRLSETS IKDRMAKYAAVSKQSSSTNYTNELKASGGGIKI 300
DB 241 LSSSAFNEKESKRNLPLRLSETS IKDRMAKYAAVSKQSSSTNYTNELKASGGGIKI 300
QY 301 HKMEQKENVPPGPEVCIITHQGEKISANENSILAVRSTPAEDDSDSOVKSEVQOPVHPKP 360
DB 241 LSSSAFNEKESKRNLPLRLSETS IKDRMAKYAAVSKQSSSTNYTNELKASGGGIKI 300
QY 299 HKMEQKENVPPGPEACSVHSGSKVSTTENSILVALSVPAEDDTCNSQVKSQAQPMHPKP 358
DB 359 HKMEQKENVPPGPEACSVHSGSKVSTTENSILVALSVPAEDDTCNSQVKSQAQPMHPKP 358
QY 361 LSPDSRASSLSSESSPPKAMKFKQAPARETCVCEQKTVYPMERLLANQOVFHSICFRCSYC 420
DB 359 LSPDARTSSLPESSPSKTAKKFQAPAKESCVCEQKTVYPMERLLANQOVFHSICFRCSYC 418
QY 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEELIERPAQ 480
DB 419 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEELIERPAQ 478
QY 481 LANARETPHSGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRIAWPPPELGG 540
DB 479 PPNAGESPHSGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRIAWPPPELGG 538
QY 541 SGSALEBEGIKMSKPKWPPPEDEISKPEVPEDVDLKLRRSSSLKERSRPTVAASFQST 600
DB 539 SGSALEBEGIKMSKPKWPPPEDEISKPEVPEDVDLKLRRSSSLKERSRPTVAASFQST 598
QY 601 SVKSPKTVSPPIRKGMSEQSEESVGG-RVAERKOVENAKSKNGNVGKTTWQNKESK 659
DB 599 SIKSPKASSPSLRKGMSEQSEESVGG-RVAERKOVENAKSKNGNVGKTTWQNKESK 653
QY 660 GETGKRSKEGHSLEMENENLVENGADSDDESNFLKQSQPQPKSLNWSFVDNTFAEEF 719
DB 654 GEEVPRSKDRSSFELESENFENGANIADDDNHVHAQQSPLEPEAPGWSGVDTTAAKEF 713
QY 720 TTQNKSDQVQWEGEVVRELVSVEEQIKRNYRDEDEEE 759
DB 714 TTQNKSDQVQWEGEVVRELVSVEEQIKRNYRDEDEEE 753

```

RESULT 2

```

Q8C7S2 ID Q8C7S2 PRELIMINARY; PRT; 753 AA.
AC Q8C7S2;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573 (2002).
DR EMBL; AK049350; BAC33699.1;
SQ SEQUENCE 753 AA; 84089 MW; 39571A17DF21F2C0 CRC64;

```

Query Match

```

Best Local Similarity 75.2%; Score 2955; DB 11; Length 753;
Matches 575; Conservative 60; Mismatches 117; Indels 8; Gaps 3;

```

```

QY 1 MESSPFRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAAEETNMKKRNTENL 60
DB 1 MESTPFRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAAEETNMKKRNPESL 60
QY 61 SQHFRKGTITVLKKWENPCLGAESHTDSLNRSSITEIRHRAHPHAEVTSASHASGAKAQ 120
DB 61 POHFRKGTITVLKKWENPCLGAESHTDSLNRSSITEIRHRAHPHAEVTSASHASGAKAQ 120
QY 121 EQIHPRSRRLSPPEALVQGRYPHIKQGEDLKORHSTESKMKMENCIGSRHEVEKSEISEN 180
DB 121 EEHTQPKRPFGRREAVIQSRYPSENSHDPKFAQATESQKMKENCIGDGRHEAEKPEISEN 180
QY 181 TDASGKIKYVNPVNLRLKMWFEKEPOTQTKILRAQSRASGRKISENSYSLDDLEIGPGQ 240
DB 181 TETSGKIEKYVNPVNLRLKMWFEKEGHNTQKSLWTSQSRNAGRRRLSENNSLDDWEIGAGH 240
QY 241 LSSSTFDESKNESRRNLPLRLSETS IKDRMAKYAAVSKQSSSTNYTNELKASGGGIKI 300
DB 241 LSSSAFNEKESKRNLPLRLSETS IKDRMAKYAAVSKQSSSTNYTNELKASGGGIKI 300
QY 301 HKMEQKENVPPGPEVCIITHQGEKISANENSILAVRSTPAEDDSDSOVKSEVQOPVHPKP 360
DB 299 HKMEQKENVPPGPEACSVHSGSKVSTTENSILVALSVPAEDDTCNSQVKSQAQPMHPKP 358
QY 361 LSPDSRASSLSSESSPPKAMKFKQAPARETCVCEQKTVYPMERLLANQOVFHSICFRCSYC 420
DB 359 LSPDARTSSLPESSPSKTAKKFQAPAKESCVCEQKTVYPMERLLANQOVFHSICFRCSYC 418
QY 421 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEELIERPAQ 480
DB 419 NNKLSLGTYSALHGRYICKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEELIERPAQ 478
QY 481 LANARETPHSGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRIAWPPPELGG 540
DB 479 PPNAGESPHSGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRIAWPPPELGG 538
QY 541 SGSALEBEGIKMSKPKWPPPEDEISKPEVPEDVDLKLRRSSSLKERSRPTVAASFQST 600
DB 539 SGSALEBEGIKMSKPKWPPPEDEISKPEVPEDVDLKLRRSSSLKERSRPTVAASFQST 598
QY 601 SVKSPKTVSPPIRKGMSEQSEESVGG-RVAERKOVENAKSKNGNVGKTTWQNKESK 659
DB 599 SIKSPKASSPSLRKGMSEQSEESVGG-RVAERKOVENAKSKNGNVGKTTWQNKESK 653
QY 660 GETGKRSKEGHSLEMENENLVENGADSDDESNFLKQSQPQPKSLNWSFVDNTFAEEF 719
DB 654 GEEVPRSKDRSSFELESENFENGANIADDDNHVHAQQSPLEPEAPGWSGVDTTAAKEF 713
QY 720 TTQNKSDQVQWEGEVVRELVSVEEQIKRNYRDEDEEE 759
DB 714 TTQNKSDQVQWEGEVVRELVSVEEQIKRNYRDEDEEE 753

```

RESULT 3

```

Q8CD09 ID Q8CD09 PRELIMINARY; PRT; 593 AA.
AC Q8CD09;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of

```


QY 161 MENCLOESRHEVEKSEISENTDASGKIEKYNVPLNLKXMPFKGEBTQTQKILRAOSRSAS 220
Db 1 MENCLOESRHEAEKPESETSETTSKIEKYNVPLNLKXMPFKGEBTQTQKILRAOSRSAS 60
QY 221 GRKISNSVSLDLEIGPGGLSSSTDSKSEKNSRRNLEPLRLSETSGIKDRMAKYQAAVSK 280
Db 61 GRKLSNNSCLDWEIAGHLSLSAENSEKNSRRNLEPLRLSETSGIKDRMAKYQAAVSK 120
QY 281 QSSSTNYTNELKASGGIEKIKHVEQENVPVPGPEVCITHQEGEKISANENSIAVSTPAE 340
Db 121 QSSPASVYNELKTS--ESKTHKVEQENVPVPGPEACSVHOGSKSVSTTENSIALSVPAE 178
QY 341 DSDSDQVSKVEQVPHKPLSPDSRASSLSSESSPPKAMKKFOAPARETCVSCQKTVYPM 400
Db 179 DDCNSQVSKSEAOQPHKPLSPDARTPLSPSSPKTAKKFOAPAKESVCEVCQKTVYPM 238
QY 401 ERLANQOVPHISCPCSYCNKLSLGTVASLHGRYCKPHFNOLPKSGNVDGFGHRP 460
Db 239 ERLANQOVPHISCPCSYCNKLSLGTVASLHGRYCKPHFNOLPKSGNVDGFGHRP 298
QY 461 HKDLWASKNEBILERRPAQANARETPHSPGVEDAPIAKVGVLAAMEAKASSQOEKED 520
Db 299 HKDLWASKNSNEETLGRPAOPNAGESPHSPGVEDAPIAKVGVLAAMEAKASSQOERED 358
QY 521 KPAETKKLRIAMPPTTELGGSGSALREGIKMSKPKWPPDEISKPEVPDVLKLLRR 580
Db 359 KPAETKKLRIAMPPEABLGSGSALREGIKVSKPKWPPDDVCKTEAPEDVLKLLRR 418
QY 581 SSKLSKRSRFTVAASFQSTSVKSPKTVPPPIKKGWMSQSESVGG-RVAERKQVENA 639
Db 419 SSKLSKRSRFTVAASFRTSSIKSPKASSPSLRKGSSEQSEFFGGGIATMERKQTENA 478
QY 640 KASKKNGVCKTTWQNKESKGTGKSKSGHSEKHEMENENLVENGADSDDDNSFLKQSP 699
Db 479 RPGEKENVCKSRQ-----GEEVPSKORSSFELESENFENGANIEDDHHVHAQQSP 533
QY 700 QBPKSLNWSFVONTFAEFTTQNKSQDVVELWEGEVKLSVEEQIKNRYYDEDEDEE 759
Db 534 LEPEAPGWSGFVDTTAAKEFTTQNKSDQVGFWEGEVWRELSVEEQIKNRYYDEDEDEE 593
RESULT 6
Q9DEY8 PRELIMINARY; PRT; 629 AA.
ID Q9DEY8 AC Q9DEY8
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Cytoskeleton-associated LIM domain protein.
GN EPLIN.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21100452; PubMed=11179679;
RA Maul R.S., Sachi Gerbin C., Chang D.D.;
RT "Characterization of mouse epithelial protein lost in neoplasm (EPLIN)
RT and comparison of mammalian and zebrafish EPLIN";
RL Gene 262:155-160(2001).
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
DR ENBL; AF307846; AAG31149.1; -.
DR HSP; P32965; 1CTL.
DR ZFIN; ZDB-GENE-001120-1; eplin.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.

DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
KW LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 629 AA; 70037 MW; C42341B024818C03 CRC64;
Query Match 28.7%; Score 1128; DB 13; Length 629;
Best Local Similarity 36.9%; Pred. No. 6.5e-59;
Matches 287; Conservative 110; Mismatches 212; Indels 168; Gaps 28;
QY 1 MESSPFNRRWTSLSLRVTAKELSLVN-KNKSSAIVFISFYKAAEETNMKKRSNTEN 59
Db 1 MVSFRRGQWASQSLRVTAKEISLVGRGKNTAIAERFSKYQAABETSLLDKKS-PEK 59
QY 60 LSQIFRKQTLVLKKKNENPGLGASHTDSLRLNSTIRHRADHPHAEVTSHAASGAKAD 119
Db 60 STPLGRNLNLVLKQLEHP---AETPT-----SPEPKAHLQNLH 96
QY 120 QEOIHPRSRLRSPEALVQGRYPHIKDGEDLKHSTESKMCNCLGESRHEVEKSEISE 179
Db 97 --QASAVKIPLESTDVQLEIG-----TD-----QOCLSDSDQPMK---RT 132
QY 180 NTDASGKIEKYNVPLNLKXMPFKGEBTQTQKILRAOSRSASGRKISENSYSLDLEIGP 239
Db 133 QRDVETLLEKPTVPLNSLKXMPFKGETLHNSVRE-----PG 169
QY 240 QLSSSTFDSKNEKNSRRNLE--LPRLSETSIKDRMAKYQAAVSKSSSTNYTNELKASGGE 297
Db 170 TTGSGSDNMPEPTKESLECGVKMLDSTPLRDRVAMYQAAVTKLDFFSSPNSE--AADSE 227
QY 298 IKIHMEQENVPV--GPEVCITHQEGEKISANENSIAVSTPAEDSDSDQVSKVEQV 356
Db 228 ARAHSGKQENVPVPSADV-----PESNTIKSP-----256
QY 357 HPKPLSPDSRASSLS-ESSPPKAMKKFOAPARETCVCEQKTVYPMERLANQOVPHISCF 415
Db 257 -----APDRNGSVLSPEQKPLVMPPLPVRETCTVCLKTVYPLEKLVANQOYHNTCF 311
QY 416 RGSYCNKLSLGTVASLHGRYCKPHFNOLPKSGNVDGFGHRPHKOLWASKNE--NEE 473
Db 312 RCAYCNTKSLVNVASLHNNYCKPHYCQLFKAGNYDEGFGHRPHKELWEGRPEGVDQ 371
QY 474 ILERPAQANARETPHSPGVEDAPIAKVGVLAAMEAKASSQOEKEDKPAETKKLRIAMP 533
Db 372 VKLSP-----QET--SLTVEESPLVKNVLAATLETRQTATSERVEKPLETGLKISWP 423
QY 534 PTELGGSGSAL---EEGKMSKPKWPPDEISKPEVPDVLKLLRRLSSSLKERSRP 590
Db 424 PQSEGESATHVTGSGIKPIRKWPPGDTVSSNV--DLESPLPKURRSVLSKERSKP 481
QY 591 FTVAASFQSTSVKSPKTV--SPPIKKGWMSQSESVGGVRAERKQVENAKAKKNGNV 648
Db 482 FCI---FDSAPVAQPKRCQSPSNEKPDSEEMSPVS-----STDTTISSED---526
QY 649 GKTWQNKESKGTGKSKSGHSEKHEMENENLVENGADSDDDNSFLKQSQPOE--PKSLN 706
Db 527 ---MTEHQSEEDQDKTKEDEQMEHSEK-----VDVQEEELSLKLCSTPDNSPPLSPE 577
QY 707 WSSFVDNTFAEFTTQNKSDQVVELWEGEVVKE-----LSVEEQIKNRYYDEDEDEE 759
Db 578 SESGLD-----PEENQASQDVGFWDGEEAEDTADVTVEDLTKRNRHYDDEDED 627
RESULT 7
Q8HXJ9 PRELIMINARY; PRT; 951 AA.
ID Q8HXJ9 AC Q8HXJ9
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.

NCBI_TaxID=9541;
[1]
SEQUENCE FROM N.A.
TISSUE=Medulla oblongata;
Hashimoto K., Osada N., Hida M., Kusuda J., Sugano S.;
Submitted (DEC-2002) to the EMBL/GenBank/DDJ databases.

[2]
SEQUENCE FROM N.A.
TISSUE=Medulla oblongata;
MEDLINE=21458551; PubMed=11574149;
Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,
Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K.;
RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
chromosomes";
Gene 275:31-37(2001).
EMBL: AB097518; BAC41743.1; --
Hypothetical protein.
KW Hypothetical protein.
QY SEQUENCE 951 AA; 107682 MW; 42A3F8DEC05274E2 CRC64;

Query Match 16.3%; Score 641; DB 6; Length 951;
Best Local Similarity 27.7%; Pred.No.8.3e-30;
Matches 212; Conservative 102; Mismatches 221; Indels 230; Gaps 28;

QY 65 KGTLTVLKXWNPNG-----LCAESHTDLSRNSSTEIRHADHPAEVTSHAASAKA 118
|||:|||::|::|:
Db 19 QKSLNLRKQWESGCVQRSECYPRDSRCTILQPQESKLL---EPPEEVVS----- 66

QY 119 DQEIQIHRPS-----RLRSPPALVOGRYPHIKDGEDLDKHSTESKKMENCLEGRHE 171
|||::|::|::|::|:
Db 67 -AREPLDPTSLPSCGETLSKP-----GKDSVDKSNTRFY-----GRPE 107

QY 172 VEKSEISENTDAGKIERYNVNLRMFKPEGEPTQTILRAQSRSASGRKIS-ENSYS 230
|||::|::|::|::|:
Db 108 VLK--EDSLSRRRIERFSIALDELRSVFAPKSGN-----KSAHGKGKEVEIERSLC 158

QY 231 LDDLEITPG-QLSSTTDS-----KNESRRNLELP 261
|||::|::|::|::|:
Db 159 SPAFKSPHQLEDVSKDSDSEKGEETSCDKVSPESHSHFEATGNPKPESPAEDSAA 218

QY 262 LSE-----TSIKDRMAKYQAIVSK--QSSSTNYTNELK--ASGEIKIHMKQKE 307
|||::|::|::|::|:
Db 219 LGVVSDLHEVSLKERMARYQAAVSGDCRFSANMWEESEMCTVGGIAKVKKQFEDE 278

QY 308 NVPPGPVCIT-----HQEGEKISANENSLAVRSTAEDDSRDQVKSEVQQ- 354
|||::|::|::|::|:
Db 279 -----ITSRNTPAQYQYQHNSREQEAHSSHQSQVTGSKSQELARNEQEGSKYQKI 329

QY 355 -----PVHPKPLSPDSR----- 366
|||::|:
Db 330 DVHGTENVSHLEKHTKEINQASOFHVQVQETVIDTPEDEEIPKYSTKLKEQFEKSQAEK 389

QY 367 --ASSLSESPPMKKFPAPARETCVECKTVYPMERLLANQOVFHISCRFCSYCNKL 424
|||::|::|::|::|:
Db 390 ILVSDKEMTTPAQIKILLQDKEICILCKTVYPMECVLADQNKFKAFCFRCHHCNSKL 449

QY 425 SLGYASLGHRICYKPHFNLFKSKGNVDEGFGRPHKDLWASKENEET----LERPAQ 480
|||::|::|::|::|:
Db 450 SLGNAYASLHGQICYKPHFKQLFKSKGNVDEGFGRGHQKHDMWNCKNQSRSDVDFPNPEPNV 509

QY 481 LANARETHPGVGVEDAPIAKGVLAASNEAKASSQOKE-DKPAETTKLIARIAMPPPTTELG 539
|||::|::|::|::|:
Db 510 CKNTAENTLVPGDNRN-----EHLDAQDSGQRNDLRKLGERGLKVWPSPKEIP 559

QY 540 SSGSALBEGIKMSKPKWPPEDISKPEVPEDVDLLKLRSSSLKERSR----- 589
|||::|::|::|::|:
Db 560 KKTTFPEEELRNKPKWPP--EMTTPUSPE-----FKSLSLLEDVTRPENKGGOGED 608

QY 590 --PFTVAASPQTSVSKSPKTVSPPIRKGSWSMQSESQSVGRVAERQVENAKASKNGN 647
|||::|::|::|::|:
Db 609 HLFF-LQPYLQSHVC-----QKEDVIG--IKEMQMHEVRKDEKKE-- 646

QY 648 VGKITWOKESKETGKRSGEHSLENENLVENGADSDSDNS 692
|||::|::|::|::|:
Db 648 VGKITWOKESKETGKRSGEHSLENENLVENGADSDSDNS 692

[illegible]

QY 441 HFNLKSKGNVDEGFGHPRKDLWASKNENBEI-----LERRAQLANARETTPHSPG 492
DB 453 HFQLFKSKGNVDEGFGHQRKDRWNCNQRSDVFPNBPENCKMCKIAENTLVP 508

RESULT 9
Q8BGB5 PRELIMINARY; PRT; 128 AA.
AC Q8BGB5
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical LIM domain/LIM domain profile/cytochrome c family
DE heme-binding site containing protein.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body, Hippocampus, and Olfactory brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
DR EMBL; AK012581; BAC25371.1; -
DR EMBL; AK032430; BAC27866.1; -
DR EMBL; AK049809; BAC33928.1; -
KW Hypothetical protein.
SQ SEQUENCE 128 AA; 14237 MW; ADF9161771331D13 CRC64;

Query Match 9.2%; Score 363; DB 11; Length 128;
Best Local Similarity 58.7%; Pred. No. 2e-14;
Matches 64; Conservative 16; Mismatches 27; Indels 2; Gaps 1;

QY 362 SPDSRASSLSSESPPKAMKFF--QAPARETCVECKTVPMERLLANQVPHISCFRCYS 419
DB 11 TPSHEAKGSSGSSVQRSKSFSLRAQVKETCAACQKTVPMERLVADKLIFHNSCFCCKH 70

QY 420 CNKLSLGTVASLHGRYCKPHFNQPKSKGNVDEGFGHPRKDLWASK 468
DB 71 CHTKLSLGSVAAMGEFYCRPHFQOLFKSKGNVDEGFGHPRKDLWASK 119

RESULT 10
Q9BT23 PRELIMINARY; PRT; 127 AA.
AC Q9BT23
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein FJ34982.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Horiba T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hoshita T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RA "NEDO human cDNA sequencing project";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
DR EMBL; BC004400; AAH04400.1; -
DR EMBL; AK092301; BAC03855.1; -
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS00478; LIM DOMAIN_1; 1.
DR PROSITE; PS00023; LIM DOMAIN_2; 1.
KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 127 AA; 14070 MW; B63174FCF0486956 CRC64;

Query Match 9.1%; Score 357; DB 4; Length 127;
Best Local Similarity 50.7%; Pred. No. 4.4e-14;
Matches 69; Conservative 16; Mismatches 27; Indels 24; Gaps 3;

QY 333 AVRSTPAEDSDSDSQVSEVQVPHKPLSPDSRASSLSSESPPKAMKKFOAPARETCVE 392
DB 7 AAQATPSHDAKGG--SSTVQ-----RSKSFS-----LRAQVKETCAA 42

QY 393 CQKTVPMERLLANQVPHISCFRCYSYCNKLSLGTVASLHGRYCKPHFNQPKSKGN 452
DB 43 CQKTVPMERLVADKLIFHNSCFCCKHCHTKLSLGSVAALHGEFYCKPHFQOLFKSKGN 102

QY 453 DEFGHPRKDLWASK 468
DB 103 DEFGHPRKDLWASK 118

RESULT 11
Q96S91 PRELIMINARY; PRT; 127 AA.
AC Q96S91;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Li N., Zhang W., Zhang M., Wan T., Cao X.;
RA Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
DR EMBL; AY037154; AAK67634.1; -
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS00478; LIM DOMAIN_1; 1.
DR PROSITE; PS00023; LIM DOMAIN_2; 1.
KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 127 AA; 14142 MW; B63014PEF0486954 CRC64;

Query Match 9.1%; Score 357; DB 4; Length 127;
Best Local Similarity 50.7%; Pred. No. 4.4e-14;
Matches 69; Conservative 16; Mismatches 27; Indels 24; Gaps 3;

QY 333 AVRSTPAEDSDSDSQVSEVQVPHKPLSPDSRASSLSSESPPKAMKKFOAPARETCVE 392
DB 7 AAQATPSHDAKGG--SSTVQ-----RSKSFS-----LRAQVKETCAA 42

QY 393 CQKTVYPMERLLANQOVFIHSCFRCSYCNKLSLGTYSALHGRYICKPHFNOLFKSGKNY 452
Db 43 CQKTVYPMERLVADKULIFHNSCFCKCHTKLSLGSVAALHGFYCKPHFQOLFKSGKNY 102
QY 453 DEQFGHRRPKDLWASK 468
Db 103 DEQFGRRQKHELWAHK 118
RESULT 12
Q9LH98 PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T198.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Columbia;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones".
RL DNA Res. 7:217-221(2000).
DR EMBL; AP002057; BAB03174.1;
SQ SEQUENCE 2081 AA; 232851 MW; D3603ELF8SEFF29 CRC64;
Query Match 6.6%; Score 258; DB 10; Length 2081;
Best Local Similarity 20.6%; Pred. No. 1.1e-06;
Matches 178; Conservative 143; Mismatches 317; Indels 226; Gaps 37;
QY 13 SLSURVAKELSLVNKSSNAIVEIPKQAAEETNNEKKRSNTENLSQHFRKGLTUL 72
Db 450 NLENKVGNEEL-----KGNASVEAKTNBESSKEEESQRSNEVYNNKTTKG----- 498
QY 73 KKKWENPGLAESHTDSLRSSTE-----IRHAD-----HPPAEVTSHAASGA 116
Db 499 -----ENVNIQGESIGDSTKNSLENKEDVKPKVDANESDGNSTKERHQBQAVNNGVSTED 554
QY 117 K-----ADQEQIHPRSLRSPPEALVQGRYPHIKD-GECLKDHSSTESKMKWNCLESR 169
Db 555 KNLNIGADEQKK-----NDKSVEVTNDGDHTKREKRETQNGESVKNENL-----E 603
QY 170 HEVKSSEISENTDAGKIEKYNVPLNKLXMMFEKPTQ-----TKILRAQSRASG 221
Db 604 NKEDKELKDDSEYGAKE-----TNNETSLEEKREQTQKHGDNINSKIVDNKGGNADS 656
QY 222 RKISE-----NSVSLDD-----LEIGPGLSSSTFSEKQNESRNLELPL-----S 263
Db 657 NKEKEVHVGSDTNDNNMESKEDTKSEVEVKKNDGSEKGEKGNKNSMEDKKLENKES 716
QY 264 ETSIKDRM-----AKYQAAVSKOSSSTNVTNKL-----A 293
Db 717 QTSKDDKSVDDKQEEAQIYGEISKDDKSVKAEAKKKESKENKTK-TNENVRNKEENV 775
QY 294 SGGSIKIHMEQKENVPPGPEVCITHOGEKISANENSLAVRSPABDSDRDSOVKSEVQ 353
Db 776 QGNKKESEKVBKESKSKDAKSVETKONKLSSTENRDEAKERSGDNKEDKE-ESKDY 834
QY 354 QVHPKPLSP-----DSRASSLSSESPPKAMKFOAPA-RETVCQKTVYPMERLLANQ 408
Db 912 QVHPKPLSP-----DSRASSLSSESPPKAMKFOAPA-RETVCQKTVYPMERLLANQ 408

Db 835 OSVEAKEKXNGVDVTVNGNKEDSKDLKDRSVEVVKANKESMKKKREEVORNDKSKSTKE 894
QY 409 VFHISCFRCSCYCN--KLSLGTYSALHGRYICKPHFNOLFKSGKNYDEGFGHRRPKDL- 464
Db 895 V-----RDFANNVDIDVQKSGSESVK---YKDE-----KKEGKEE-----NKDTI 933
QY 465 -WASKNENEIEILRPAQLANARETPHSPGVEDAPIAKYGVLAASWEAKASQOE----- 517
Db 934 NTSSKQKQKDKKKKXKESKN-----SNMKKKEEDKKEYVNNEL 971
QY 518 --KEDKPAETKKLRIAMPPTTELSSGSGALBEGIKMSKPKWPPPEDEISK-----PEVP 568
Db 972 KQEDNKKETTK-----SENSKLKEENKDKKEKSEDSASKNREKKEYEKK 1019
QY 569 EDVDLDLKLRRSSSLKRSRPFVVAASFQSTSVKSPKTVPPPIKGMMSFSQSESVGG 628
Db 1020 SRTKEAKKEKKKSQDKKREE-----KQSEERKSKKEEESRDLK-----AKKKEE---- 1065
QY 629 RVAERKQVENAKASKNGNVGKTTWQNKESKGETKGRSKGHSLEMEENLVENGADSD- 687
Db 1065 ETKKESKESNHSKKKEDY--KEHEDNKSMMKEEDKKEKKKHE-ESKRRKKEEDKKDMEK 1122
QY 688 -EDNNSFLKQSQPBPKSLNWSFYDNTFAEFTTQNG-----KQSDVELWEGE 735
Db 1123 LEDQNSNKKKDKNEKKKSQHVKLKESDKKEKKEBEKSETKIESSKSQKNEVDKKE 1182
QY 736 VVKELSVBEQIKRNYDEDEDEE 759
Db 1183 --KXSKDQKQKKEKEMKESEBK 1204
RESULT 13
Q9W596 PRELIMINARY; PRT; 5412 AA.
AC Q9W596;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE CG3064 protein.
GN FUTSCH OR EG:494.1 OR CG3064.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Barnkocho C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Jalaali M., Kaulush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskeirn D.R., Pacht J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celnik S.E., Adams J.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne M.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernier B., Carlson J.W., Celnik S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003420; AAF45622.3; ..
DR FlyBase; FBgn0015390; futsch.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR000531; TonB_boxC.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
SQ SEQUENCE 5412 AA; 584513 MW; 7881CAAC8749FAFF CRC64;

Query Match 6.4%; Score 251; DB 5; Length 5412;
Best Local Similarity 20.7%; Pred. No. 1e-05;
Matches 199; Conservative 138; Mismatches 359; Indels 264; Gaps 37;

QY 22 ELSLVNKNSSAIVFISKYQKAAETNMKRSNTNLSOHFRKGLTVLKKWNPGL 81
DB 954 EDSIVQESSMTKEEIQHQDSQSEKKKKAEEIEAIAKVAEAKRALE---- 1009
QY 82 CAESHTDSLRSSTIRRHADHPPAVTSHAAGS---AKAQEEQI--HPRSRLRSP-PE 135
DB 1010 GASARQD---ESELDEVEPQSKIAEVQDIATAKDIKRSIEQLAKPAEELSSTPE 1066
QY 136 ALVQGRYPHIKDG-----DLKHSTESK-----X 160
DB 1067 EKLSKKTSQDDQIGAPVDLPVNLQESLPEERKFSATIESGATTAPLTPEDERIPLDQI 1126

QY 161 MENCIGSRHEVEKSEISE-----NTDASGI 187
DB 1127 KEDLVIEEYKVEETKEAEIAVVAIVQTLPEAAPLAIDTILASATKDAPKANAAALGEL 1186
QY 188 EKYNVPLNRLKNMFKGEPQTQILR-----AQSRSASGRK 223
DB 1187 PDSGERVLPKMTFE-----AQQLLRDVIKTPDEVADLPVHEEADGLYKDSQDAGAKS 1242
QY 224 ISENSYSL-----DDEIGFGQLSSSTDFSEKNESSRNLEPLRLSETSIKDRVAKYQA 276
DB 1243 ISHKEESAEEKETDDEKENKVGTE---LGDEPNKVDISHVLLKESVQVAAKVVVIET 1299
QY 277 AVSKQ-----SSSTNYTHLKAAGGEIKHMEQENVPVPGPEVCIHQEGEKISANENS 331
DB 1300 TVEKQEEIVEATTITOENQEDLMEQVKDEEHEQKTESG---ITKEAKKSA----- 1351
QY 332 LAVRSTPAS-----DSDRSOVKSEVQVQHPKPLSPDGRASSLSSESP-----K 377
DB 1352 -----STPEKETSDITSDDELPQAADPTTPPKSAKREDTGSIESPTIEAEIIEVQ 1407
QY 378 AMKKFO--APARETCVECOCTYPMERLIANQQVFHISCFRCSYCNKLSLGTYSALHGR 435
DB 1408 AKQEAQKVPAPAEAIKTEKSP-----LASKETSRPESATGSV---KEDTEQTKSKSP 1458
QY 436 IYCKPHFNOLFKSKNGYDEFGHRP-----HKDLWASKNENEELILERP 479
DB 1459 VPSRPE-SEAKDKSPFASGEASRPESVAESVYKDEAGKAESRRSIAKTHKDESLDKAK 1517
QY 480 QLANARETHSPQVEDAPIAKVGLAASMEAKASSQOEKDP-----AETKKL 528
DB 1518 EQESRRRESIAESIKPESGIDEXSALASKEASRPESVTDKSKPESRESIAESLKAESTKD 1577
QY 529 RIAMPPTTELGGSSGSALE-----EGIKMSKPK-----WPP-----E 559
DB 1578 EKSAAPSKASRPGSVVSVKDETEKSPRESIAESAKPIEPREYSRRESVIDGIK 1637
QY 560 DEISKPEVPEDVDLKLRRSSSL-----KERSRPTVAASFQSTSVKSPKTV 608
DB 1638 DESAKPESRRDGPLASKEASRPESVLESVKDEPIKSTKESRRESVAESFKADSTKDEK-- 1695
QY 609 SPIRKGNMSQSESVGGRV--AERKQVENAKSKNGNVTGKTWQNKESKGETGKRS 666
DB 1696 SPLTSKDISRPESAVENVMVDAVGSARSQSPESVTASR---DVSRPESVAESKDDTKDPE 1752
QY 667 KECHSLEMENEL-VENGADSDDEDNSFL-----KQSPQE-----PKSLNW 707
DB 1753 SVVESVIPASDVVEIEKGA-ADKEKGVFVLSIGKPDSPSEVISRPGVPVSVKESRRE 1811
QY 708 SS--FVDNTFAEEFTTQNKQSDVELW--EGEVVK-----ELSVEEQIKENRYIYDEDEEE 759
DB 1812 SSTEIVLPCHAEDSKESRPESKVECLKDESEVLKSGSTRRESVAESDKSSQPFKTSRPE 1871
RESULT 14
ID 076891 PRELIMINARY; PRT; 5327 AA.
AC 076891;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE EG:49E4.1 protein.
GN FUTSCH OR EG:49E4.1 OR CG3064.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Papagiannakis G., Spanos L., Siden-Kiamos I., Louis C.;
RA "Sequencing the distal X chromosome of Drosophila melanogaster";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

RN (2)
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031128; CAA20006.1;
DR FlyBase; FBgn0015390; futsch.
SQ SEQUENCE 5327 AA; 575942 MW; FEFEE23A118FF38A CRC64;

Query Match
Best Local Similarity 22.3%; Pred. No. 1.9e-05;
Matches 192; Conservative 124; Mismatches 330; Indels 214; Gaps 38;

QY 77 ENPGLCAESHTSLNSSTFIRHRAHPAETVSHAASAKADQEQIHPRSLRSPPEA 136
DB 2600 ERPSPAESAADA--ESVEKSKDASRPSPSVVSTKADSTKDISPS--PSVLEGPDKD 2655

QY 137 LVQ-----GRYPHIK---DGEDLKHS-----TESKMKENCLGESRHE-----VE----- 173
DB 2656 VEKSKESRRPPVSASITGDSKDVSRPASVSVKDEHDAESRRSIAKVESVIDEAG 2715

QY 174 KSEISNTDASIKIKYV--PLNRLKMFKEGPTQTKI-LRAQSRASGRKISENSY 229
DB 2716 KGDSSSSQDSQDKSTILASKEASRRSRSVWSSKDDAEKSSRPESVIAAGPEVPRESK 2775

QY 230 S-LDLEIG-PQG-LSSSTFDSEKNE--SRNLELPRLSETSIKDRMAKYQAA----- 277
DB 2776 SPLDSKOTSRPGSVSVTAEDKSEQSRRESVAESVKAADTKD--GKSOEASRPSSVD 2833

QY 278 -----VSKQSSTNYTNELKASGEIKIHKMEQENVPVPGVCITHQGEKISA 327
DB 2834 ELLKODDEKQESRRSITGSHKAMSTMGDESMPDKAD-KSEKSPSPESVASKIKHENTKD 2892

QY 328 NENSLAVSTPAEDSDRSQVKEV-----QQVHPKPLS-----PDSRASSL 370
DB 2893 EESPLGSR-----DSVAESIKSDITKGKSLPSKEVSRPVSIGSIKDEKAESRRSV 2947

QY 371 SSSSPKAMKQA--PARETC-----VECKTVYPMERLLANQ 408
DB 2948 AESVPESSKDATAPPSKEHRSRPESVGLSKDEGDKTTSRVVADSIKDEKSLLSQ 3007

QY 409 VFHISFCRCSYCNKLSTGYASLHGRYCKPH-----FNQLFK 447
DB 3008 -----ASRSEASLKDAAASQETSRPESVTSVKDGKSPVASKESRPRASVAEN 3059

QY 448 SKGNVDEGHRP-----HKDLWASKHNEEILERPAAQLANARETPHSGVE 494
DB 3060 AKDSADESKQRPESILPQSKAGSIKDEKSLASKDEAKSKEESRRESVAEQFP-----LV 3115

QY 495 DAPIAKGVLAAMEAKASSQEKDKPAETKKLRIAMPPTTELSSGSALEEGIKMSK- 553
DB 3116 SKEVSRPASVAESV--KDEAKSKESPLMSKE--ASRPASVAGSVKDEAKSKESRR 3170

QY 554 ----PKWP-PEDEISKP-----EVPEDVDLDLKKLRSS-----SLKERSRPTVAAS 596
DB 3171 ESVAESPLPSKEARPASVAESVKDEADKSESRRESGAESKPLASKEASRPASVAES 3230

QY 597 F-----OSTSVKSP---KTVSPPIRKGMMSSEQS-----EESVGRVRAERQV 636
DB 3231 KDEAKSKESRRRESVAESKPLPSKEASRPSTSVASKSVKDEAKSKESRRSDSVAESKPL 3290

QY 637 ENAKASKNGVYKTTWQNKESGTGRSK--EGHSLMEENENILVNGADSDENSL 694
DB 3291 ASKEASRP-ASVAESQDEAKSKESRRRESVAESKPLAYKASRPASVAESIKDEAKS 3349

QY 695 KQCSQOE-----PKSLNWSFVDNTFAEFTTQNKQSDQVELWEGEVYKE 739
DB 3350 KEESRRRESVAESKPLASKEASRPSTSVASKSVKDEAKSKESRRSDSVAESKPLASKEASRP 3409

QY 740 LSVEEQIKNRYDDEDEE 759
DB 3410 ASVAESV-----QDEAK 3422

RESULT 15
Q9V7G8 PRELIMINARY; PRT; 7210 AA.
AC Q9V7G8;
AD 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG18255 protein.
GN STRN-MLCK OR CG8304 OR CG18255.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jallali M., Kalush H., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacלב J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
(2)
RP SEQUENCE FROM N.A.
RA Celnik S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwan C., Jallali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacלב J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[3]
SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,
RA Clamp W., Drysdale R., Bamert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Seasholtz S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003908; AAF58087.2; -
DR FlyBase; FBgn0013988; Strn-Mlck.
DR InterPro; IPR0031599; Ig.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 13.
DR SMART; SM00409; Ig; 21.
DR SMART; SM00408; IGC2; 20.
DR PROSITE; PS00835; IG-LIKE; 20.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 7210 AA; 808965 MW; 6B2E7395C07140D2 CRC64;

Query Match 6.1%; Score 241.5; DB 5; Length 7210;
Best Local Similarity 20.4%; Pred. No. 5.3e-05;
Matches 167; Conservative 155; Mismatches 357; Indels 139; Gaps 31;

QY 19 TAKELSLVNNKSSAIVEIFSKYQAAETNNKESNTENLSQHPKGTTLVLKKWEN 78
DB 3765 TOTDSAIDEKSAEYSEIVS--EKITDEKAEQSEKVEVD-SEAKPKKAKVLEKSEIE 3821

QY 79 PGL--GAESHTDSL---RNSSTEIRHADHPAETVSHAASGAKADQEQIHPRLRSP 133
DB 3822 EKLEDKKEKQTESAIDEKSAEYSEIVS--SEIVSEKITDEKAEQSEKVEKVGSEAKP 3873

QY 134 PEALVQGRYPHIKDGEDLDKDHSTESKMNCLGESRHEVEKSEI-----S 178
DB 3874 KKAKVLEK-----KSIEEEKLEDKKEKQTESAIDEKSAEYSEIVSEKITDEKAEQSE 3929

QY 179 ENTDSAGKIEKNVPLNRLNMMEKGEPTQTILRAQSRASASGRKISENSYSLDDLEIGP 238
DB 3930 EVKDSKPKKAKV-LEKKSIEEEKLENKKEK-----QTESA----IDEKSAEYSEI-- 3978

QY 239 GQLSSSTFTDSEKNESRRN-----LELPRLSSETSIKORMAKY-QAAVSKOSS 283
DB 3979 --VSEKITDEKAEQSEKVEKVDSEAKPKKAKVLEKKSIEEEKLEDKKEKQTESAIDEK 4036

QY 284 -----STNYTNELKASGGEIKIHKMEQKENVPPGPEVCITHOGEKISANENSLAVR 335
DB 4037 KAEVSEIVSENIITDE-KAEQSEKX--EVKDSKPKKAKV-----EKKSTIEEEKLEDK 4087

QY 336 STPAEDDSRDSQ-VKSEVOQVHPKPLSPDSRASSL-----SESSPPKMKYFQAPARET 389
DB 4088 KETQTSADAIDEKSAEYSEIVSEKITDEKAEQSEKVEKVDSEAKPKKAKVLEKKSIEEE 4147

QY 390 CVCEQKTVYPMERLLANQQVFHISCFRCSCYNNKLSLGTYSILHGRIVCKPHFNOLFKSK 449
DB 4148 KLEDDKKEKQTESAIDEKSAEYSEIVSENIITDEKAEQSEKVEKVDSEAKPKKAKVLEK 4207

QY 450 GNYDEGFGHRPHKDLWASKNE-----NEETILERPAQLANARETPHSPG--VEDA 496
DB 4208 SIEEEKLEDKKEKQTESAIDEKSAEYSEIVSEKITDEKAEQSEKVEKVDSEAKPKKAK 4267

QY 497 PIKVGVLAAAMEAKASSQOEK--EDKPAETKKLRIAMPPTTELGGSSGSALEBGIKMSKP 554
DB 4268 VLEKKSIEEEKLEDKKEKQTESAIDEKSAEYSEIVSEKITDEKAEQSEKVE--VKDSEA 4326

QY 555 KWPEDSISKEPEVPEDVDLDLKKLRRS--SSLKERSRFTTVAASFQSTSVKSPKTVSPPIR 613
DB 4327 K-PKAKVLEKKSIEEEKLENKKEKQTESAIDEKSAEYSEIVSEKIT 4374

QY 614 KGWMSQSQSEESVG-----GRVAERKQVENAKASKKNGNVGKTTWQNKESKGETGKRS 666
DB 4375 DEKAEQSEKVEKVGSEAKPKKAKVLEKKSIEEEKLEDK-----KEKQTESAIDEKS 4425

QY 667 KEGHSLMENENLVENGADSDSDNSFLKQOSQOEKSLNWSFVDNTFAEEFTTQ----- 722
DB 4426 QKAEVSEIVSEKITDEKAEQSEKVEKVDSEAKPKKAKVLEKKSIEEEKLEDKKEKQTESA 4485

QY 723 -NQASQDVLEWEGEVVKELSVEEQIKRNYDEDEEE 759
DB 4486 IDEKSAEYSEIVSEKITDEKAEQSEKVEKVDSEAKPKKAKVLEKKSIEEEKLEDKKEKQTESA 4520

Search completed: January 6, 2004, 09:49:54
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:45:49 ; Search time 21 Seconds
(without alignments)
1529.235 Million cell updates/sec

Title: US-09-890-549-4
Perfect score: 3927
Sequence: 1 MESSFNRRQWTSLSRVTA.....LSVBEQIKRNYDEDEDEE 759

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/aa/5A-COMB.pep:*
2: /cgn2_6/prodata/1/aa/5B-COMB.pep:*
3: /cgn2_6/prodata/1/aa/6A-COMB.pep:*
4: /cgn2_6/prodata/1/aa/6B-COMB.pep:*
5: /cgn2_6/prodata/1/aa/PCTUS-COMB.pep:*
6: /cgn2_6/prodata/1/aa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223	5.7	1312	4	US-09-345-882-29
2	217	5.5	219	3	US-08-934-627B-6
3	210.5	5.4	200	4	US-09-282-146-2
4	205.5	5.2	208	3	US-08-934-627B-2
5	193	4.9	2842	1	US-07-741-940-7
6	193	4.9	2842	1	US-08-289-548A-7
7	193	4.9	2842	1	US-08-452-654-7
8	193	4.9	2842	4	US-08-449-731-7
9	193	4.9	2843	1	US-07-741-940-2
10	193	4.9	2843	1	US-08-289-548A-2
11	193	4.9	2843	1	US-08-452-654-2
12	193	4.9	2843	1	US-08-452-655B-2
13	193	4.9	2843	1	US-08-452-655B-7
14	193	4.9	2843	2	US-08-370-235A-2
15	193	4.9	2843	3	US-08-450-582-2
16	193	4.9	2843	3	US-08-450-582-7
17	193	4.9	2843	4	US-08-449-731-2
18	193	4.9	2843	2	US-08-821-353A-7
19	193	4.9	2843	2	US-09-003-687A-7
20	193	4.9	2843	1	US-09-136-605-7
21	192	4.9	1780	1	US-08-769-309A-5
22	192	4.9	1780	2	US-08-994-570-5
23	190	4.8	1864	2	US-08-790-912-3
24	190	4.8	2052	2	US-08-790-912-2
25	187.5	4.8	3969	3	US-08-061-376-5
26	184.5	4.7	1939	4	US-09-310-187A-1
27	178.5	4.5	1848	3	US-08-296-791-6

ALIGNMENTS

RESULT 1

US-09-345-882-29
; Sequence 29, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; FILE REFERENCE: AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
; CURRENT APPLICATION NUMBER: US/09/345.882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 29
; LENGTH: 1312
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 294..296
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 432..434
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 755..757
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 856..858
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 859..861
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 910..912
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 1151..1153
; OTHER INFORMATION: potential
; FEATURE:

Sequence 6, Appli
Sequence 15, Appl
Sequence 15, Appl
Sequence 15, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 1, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 32, Appl
Sequence 7, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 11, Appl
Sequence 16, Appl
Sequence 2, Appli
Sequence 14, Appl

28 178.5 4.5 1848 5 PCT-US95-10661A-6
29 178 4.5 1269 4 US-09-645-456A-15
30 178 4.5 1269 4 US-09-425-324A-15
31 178 4.5 1269 4 US-09-645-791-15
32 177.5 4.5 1324 4 US-09-645-456A-13
33 177.5 4.5 1324 4 US-09-425-324A-13
34 177.5 4.5 1324 4 US-09-645-791-13
35 177.5 4.5 3248 1 US-08-353-700-1
36 177.5 4.5 3248 5 PCT-US95-16216-1
37 177 4.5 1805 1 US-07-853-913-2
38 177 4.5 8991 4 US-08-714-741-32
39 176.5 4.5 1400 1 US-08-080-255-7
40 176.5 4.5 1400 3 US-08-455-713-7
41 176.5 4.5 1400 5 PCT-US93-05857-7
42 174.5 4.4 1588 5 PCT-US93-07261-11
43 174.5 4.4 1663 5 PCT-US93-07261-16
44 174 4.4 1354 3 US-08-685-871-2
45 173.5 4.4 1298 4 US-09-645-456A-14


```
NAME/KEY: CARBOHYD
LOCATION: 1226..1228
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 102..105
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 663..666
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 808..811
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 885..888
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 17..19
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 31..33
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 41..43
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 100..102
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 140..142
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 216..218
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 471..473
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 507..509
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 531..533
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 591..593
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 656..658
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 801..803
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 812..814
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 815..817
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 876..878
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 888..890
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 939..941
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1060..1062
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1128..1130
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1129..1131
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1135..1137
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1181..1183
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1208..1210
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1249..1251
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 47..50
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 126..129
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 157..160
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 158..161
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 159..162
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 216..219
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 274..277
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 276..279
```

```

; OTHER INFORMATION: potential
;
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 295..298
;
; OTHER INFORMATION: potential
;
; FEATURE:

```

```

Query Match      5.7%; Score 223; DB 4; Length 1312;
Best Local Similarity 19.4%; Pred. No. 2.2e-09;
Matches 177; Conservative 150; Mismatches 329; Indels 258; Gaps 42;

QY 40 KYQAAEETNWKRRNTENLSQHRKGLTLVLKKWENPGLGAESHTDSLRRNSFEIRH 99
Db 403 EFQMALEPKVNVKQCECNVK-----IKVKEENETEIKIKWEERN-----IIP 449
QY 100 RADHPAEVTTSHAASAKADQEEQIHP-----RSRLRSPPEALVQGRYPHIK---DGEDL 151
Db 450 REEKPIED-----EIERKENIKPSLGSKKNLLESIPTHSDQEKVNIKPEDNENL 500
QY 152 KDHSTSKKWNENCL-----GSRHEVEKSEISENTDAS 184
Db 501 DKDDDDTTRVDESINIKVEAESEKAKSGDNTNKEEDDDEAESEEEEEDEDDDDN 560
QY 185 GKIEKYNVPLNRLKMFKEQPTQTKILRAQSAS-----CRKISENSYSL 231
Db 561 NEEEFECYPPGMKVQVRYGRGNQKNMYEASIKDSVVEGGEVLYLVHYCGMNVRYDEWIK 620
QY 232 DDEIEGPGQLS-----SSTFQSEKNE-----SRNLELPRLSETSJK-----DRMA 272
Db 621 ADKIVPADKQVVKIKHRKKIKNKLDEKDKDEKYSFQCKLRLSKPPQTNPSPMWS 680
QY 273 KYQAAVSKQSSS-----TNTNELKAGGGEIKIHMQEKENVPPGPEVCITHOGEK 324
Db 681 KLDLTDAKNSDTAHIKSIEITSINGLOQSAESAEQEDERGAQMDMNGKEESKIDH 740
QY 325 ISANENSLAVRSTPAEDSDSDSQVSE-----VQOPVHPKP-----LSPDS-- 365
Db 741 LTNRRNDLI-----SKBEQNSSLLEENKYHADLVISKPVKSPERLRKDIEVLSEDTDY 795
QY 366 -----RASSLGESSPPKAMK-KFOAPARETC--VECOKTVTP--MERLLANQOVFHS 413
Db 796 EDEVTYKRDVKDVTDDKSKSQIKRGKRRYNTBECLTGSFGKKEEKAKNKE----- 850
QY 414 CFRCSYC-----NNKLSLGTVASLHGRYCKPFHN 443
Db 851 -----SLCMENSSNSDDEETKAKMTPTKYNGLEEKRLSLTTCFYSG-----FS 899
QY 444 QL-----FKSGNYDEBFGH---RPHKDLNAS-----KNENBILERPAAOLANARTPHS 490
Db 900 EVAEKRIKLNNSDERLQNSRAKDRXDVWSSIQGQWPKTKLFLFS--DSDTEAAASPPH- 957
QY 491 PGVEDA-----PIAKGVGLAASMEAKASSQOEKDKPAETKKLRI-----AWP----- 533
Db 958 PAPEXVAESXQTVAEESCSFVLEKPPPVNVVDSKPIEKTVEVNDRKAEFFSGSN 1017
QY 534 -----PPT--ELGSS-----GSALLEGIKMSKPKWPPDEIS--KPEVPEDVDLD----- 574
Db 1018 SVLNTPTPTTPESSVTVTGSPQQSSVTVSEPLAPNQEVRISIKETOSTIEVDSVAGE 1077
QY 575 LKLRSSLSKBSRPPTVAASQOSTSVKSPKTVSPPIRKGWSMSQOSEESVQGRVAERK 634
Db 1078 LODLOQEGN-----SSPAGEDASVSSSSSNQPEPEHP--EKACTGQKRVKDAQGGSSSK 1131
QY 635 OVENAKA---SKKNGNVGKTTWQNKESKGETCKRSK-----EGHSL--MENENL 679
Db 1132 QKRSHKATVNNKKKGKNTSSDSBELSAGESITKQPVKSVSTGMKSHSTSPARTQSP 1191
QY 680 VENGASDEDDNSFLKQQSQPEPKSLNWS---SFVDN-TFAEBFTTQNKSQSDVELWEGE 735
Db 1192 GKQKNGDKDPD--LKEPSNRLPKVYKNSFQNSDLENWTSABRITILQEKLQIRKHYS 1249
QY 736 VKVEL-SVEEQIKR 748
Db 736 : : : : :

```

```

Db          1250 LKSEVASIDRRKR 1263

RESULT 2
US-08-934-627B-6
/ Sequence 6, Application US/0893
/ Patent No. 6169174
/ GENERAL INFORMATION:
/ APPLICANT: OSAMU HASEGAWA
/ APPLICANT: SATOSHI AOTSUKA
/ APPLICANT: SOICHIRO TAKENAKA
/ APPLICANT: HIROFUMI UCHIMYI
/ TITLE OF INVENTION: COTTON
/ NUMBER OF SEQUENCES: 6
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Jones & Askew,
/ STREET: 3424 Peachtree Road
/ CITY: Atlanta
/ STATE: Georgia
/ COUNTRY: USA
/ ZIP: 30326
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatib
/ OPERATING SYSTEM: PC-DOS/
/ SOFTWARE: PatentIn
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Roger T. Frost
/ REGISTRATION NUMBER: 22,1
/ REFERENCE/DOCKET NUMBER:
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (404) 949-2400
/ TELEFAX: (404) 949-2499
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 219
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-934-627B-6

Query Match          5.5%
Best Local Similarity 35.8%
Matches 48; Conservative 2

QY          379 MKKFGQAPARETCVEQKTV
Db          1 MKSFTGTTQK-CTVCEKTV
QY          439 KPHFNQLFKSGNYDEGFGH
Db          60 RHHDQLFKRTGSLKSFDS
QY          497 -----PIAKGVV 503
Db          112 ACAKIVYPIRWV 125

```

Query Match 5.5%; Score 217; DB 3; Length 219;
Best Local Similarity 35.8%; Pred. NO. 4.8e-10;
Matches 48: Conservative 23; Mismatches 45; Indels

[illegible]

RESULT 3
US-08-282-146-2
; Sequence 2, Application US/09282146A
; Patent No. 6303847
; GENERAL INFORMATION:
; APPLICANT: KAWAOKA, Akiyoshi
; APPLICANT: EBINUMA, Hiroyasu
; TITLE OF INVENTION: TRANSCRIPTION ERROR DETECTION

```

; TITLE OF INVENTION: BIOSYNTHESIS PATHWAY
; FILE REFERENCE: 4859-0027-0 US/09/282,146A
; CURRENT APPLICATION NUMBER: US/09/282,146A
; CURRENT FILING DATE: 1999-03-31
; EARLIER FILING DATE: 1999-03-31
; EARLIER FILING DATE: 1998-03-31
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Nicotiana tabacum
; US-09-282-146-2

Query Match
Best Local Similarity 5.4%; Score 210.5; DB 4; Length 200;
Matches 39; Conservative 26; Mismatches 41; Indels 21; Gaps 2;

QY 384 APARETCVECKTVYPMERLLANQQVPHISCFRCSCYNNKLSLGTYSALHGRYICKPHEN 443
Db 4 AGTTQKMACDKTVYLVLDKLTADNRVYHKAFCRHHCKGTGVLKLNYSNPFGLVYCRPHFD 63

QY 444 QLFKSKGNVDEGF-----GHRPHKDLWASKNENEIILER-----PAQLA 482
Db 64 QLFKQIGSLDKSPGTPKXVNVKPKPIDSEKPVAKVTSMPGGTREKCFGCKKTVYITEKV 123

QY 483 NARETTPH 489
Db 124 SANGTPY 130

RESULT 4
US-08-934-627B-2
; Sequence 2, Application US/08934627B
; Patent No. 6169174
; GENERAL INFORMATION:
; APPLICANT: OSAMU HASIGAWA
; APPLICANT: SATOSHI AOTSUKA
; APPLICANT: SOICHIRO TAKENISHI
; APPLICANT: HIROFUMI UCHIMURA
; TITLE OF INVENTION: COTTON PLANT GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew, LLP
; STREET: 3424 Peachtree Road, N.E., 2400 Monarch Tower
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30326
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,627B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Roger T. Frost
; REGISTRATION NUMBER: 22,176
; REFERENCE/DOCKET NUMBER: 20111-0010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 949-2400
; TELEFAX: (404) 949-2499
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 208
; TYPE: amino acid
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
; US-08-934-627B-2

Query Match
Best Local Similarity 5.2%; Score 205.5; DB 3; Length 208;
Matches 44; Conservative 21; Mismatches 59; Indels 1; Gaps 1;

QY 379 MKKFOAPARETCVECKTVYPMERLLANQQVPHISCFRCSCYNNKLSLGTYSALHGRYIC 438
Db 1 MATFQG-TQKCNACNKTIVYLVLDKLTADNRVYHKAFCRHHCKGTGTLKLSNYSNPFGLVYC 59

QY 439 KPHFNQLEKSKGNVDEGFHGHKDLWASKNENEIILERPAQLANARETTPHSPGVDEAPI 498
Db 60 RHYDQLFRTGSLDKSPGTPKVVKPERQIDSSALKVMNSFGGTREKCAACSKTAYPI 119

QY 499 AKGVV 503
Db 120 ERVTV 124

RESULT 5
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/741,940
; APPLICATION NUMBER: 19920109
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; US-07-741-940-7

```

```
Query Match      4.9%; Score 193; DB 1; Length 2842;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKNKSAIVEIFSKYKAAEET---NMKKRSNTENLSQHFR----- 65
DB 1788 VRKNAD-----SKNLNAERVSFNDKSKQNLKNNKDFNDKI PNNEDRVGSAFD 1840
QY 66 -----KGLTVLKKWENPGLGAESHTDSLNRNSTEIRHRADHPAE--VTSH---A 112
DB 1841 SPHHYPIEGTPYCFGRNDSLSLDDDDVDLSREKAEELRKAKENKESAKVTSHTELT 1900
QY 113 ASGAKADQEQIHRPSRLSPPEALVQGRYPHDKGEDLKD--STESKKMEN----- 163
DB 1901 SNOQSANKTQAIKQPINRGQPKPILOKQSTFPQSSKDI PDGAAATDEKLQFAIENTPV 1960
QY 164 CLGESRHEVEKSEISNTDASGIEKYVPLNRLKMFKEGTEPTQTKILRAQSRASGRK 223
DB 1961 CFHNSLSLSLSDIQENN-----NKNEPIKETEPDPSQGEPSKQP----- 2005
QY 224 ISENSYSLLDLEI---GPGQLSSSTFDSEKNESRRNLELPRLSSETS IKDRMAKYQAAVSK 280
DB 2006 YAPKSFHVEDTPVCFGRNDSLSLSLSDSEDD-----LQECISSAMPK----- 2050
QY 281 QSSSTNYTNEL---KASGG-----EIK-IHKMEQKENVPPGPEVCITHQEGEKISAN 328
DB 2051 KPSRLKGDNEKSPRNMGILGEDLTLDLKDIPRDPSEHGLSPDS-----NFDWKAIOEG 2106
QY 329 ENSL-----AVRSTPAEDSDS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
DB 2107 ANSIVSSLHCAAAAACLSQAASDS--DSLILSKSGISLGSPPH---LTPDQEKPTSNK 2162
QY 375 PPKAMKKFOAPARETCVQCKTVYPMERLLANQQVPHISCFRCSCYNNKLSLGTVYASLHG 434
DB 2163 GPRILK---PGEKSTLETK----- 2179
QY 435 RIYCKPHFNOLFKSKNYDEGGHRRPHKDLWASK-NENBEI---LERPAQ-----LANAR 485
DB 2180 -----IESESKGIRG-GKVKYKSLITGKVRNSEISGQMKQPLQANMPSISGR 2227
QY 486 ETPHSPGVED-----APIAKVGVLAASMEAKASSQOE-----KEDKPAETKKLRIANPP 534
DB 2228 TWIHIPGVNSSSSTSPVSKGPPLKTPASKSPSEGTATTPRGAKSVKSELSPVARQ 2287
QY 535 PTELGGSSGA-LEEGIKMSKWPPEDEISKPEVPEDVDLDLKLRLRSSSLKERS--RPF 591
DB 2288 TSQIGSSKAPSRSGSRDTPSRPAQOPLSRP-----IQSPGRNSISPRNGISPP 2338
QY 592 TAAASFQSTSVKSPKTVSPPIKAGMSMQSEESVGGVRAERKQVENAKSKNGNVGKT 651
DB 2339 NXLSQLPRTS--SPSTAS---TKSSGSGKVSYSFQRMQSQOQLTKQTGLSKNASSIPRS 2393
QY 652 TWQNKESKGTGKRSKEGHSLEMEENL---VENGADSDSDNSFLKQOS 698
DB 2394 ---ESASKGLNQMNNGANKKVELSRMSTKSSGSESDRSPVLRQS 2440

RESULT 6
US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
```

```
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; US-08-289-548A-7

Query Match      4.9%; Score 193; DB 1; Length 2842;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKNKSAIVEIFSKYKAAEET---NMKKRSNTENLSQHFR----- 65
DB 1788 VRKNAD-----SKNLNAERVSFNDKSKQNLKNNKDFNDKI PNNEDRVGSAFD 1840
QY 66 -----KGLTVLKKWENPGLGAESHTDSLNRNSTEIRHRADHPAE--VTSH---A 112
DB 1841 SPHHYPIEGTPYCFGRNDSLSLDDDDVDLSREKAEELRKAKENKESAKVTSHTELT 1900
QY 113 ASGAKADQEQIHRPSRLSPPEALVQGRYPHDKGEDLKD--STESKKMEN----- 163
DB 1901 SNOQSANKTQAIKQPINRGQPKPILOKQSTFPQSSKDI PDGAAATDEKLQFAIENTPV 1960
QY 164 CLGESRHEVEKSEISNTDASGIEKYVPLNRLKMFKEGTEPTQTKILRAQSRASGRK 223
DB 1961 CFHNSLSLSLSDIQENN-----NKNEPIKETEPDPSQGEPSKQP----- 2005
QY 224 ISENSYSLLDLEI---GPGQLSSSTFDSEKNESRRNLELPRLSSETS IKDRMAKYQAAVSK 280
DB 2006 YAPKSFHVEDTPVCFGRNDSLSLSLSDSEDD-----LQECISSAMPK----- 2050
QY 281 QSSSTNYTNEL---KASGG-----EIK-IHKMEQKENVPPGPEVCITHQEGEKISAN 328
DB 2051 KPSRLKGDNEKSPRNMGILGEDLTLDLKDIPRDPSEHGLSPDS-----NFDWKAIOEG 2106
QY 329 ENSL-----AVRSTPAEDSDS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
DB 2107 ANSIVSSLHCAAAAACLSQAASDS--DSLILSKSGISLGSPPH---LTPDQEKPTSNK 2162
QY 375 PPKAMKKFOAPARETCVQCKTVYPMERLLANQQVPHISCFRCSCYNNKLSLGTVYASLHG 434
DB 2163 GPRILK---PGEKSTLETK----- 2179
QY 435 RIYCKPHFNOLFKSKNYDEGGHRRPHKDLWASK-NENBEI---LERPAQ-----LANAR 485
DB 2180 -----IESESKGIRG-GKVKYKSLITGKVRNSEISGQMKQPLQANMPSISGR 2227
QY 486 ETPHSPGVED-----APIAKVGVLAASMEAKASSQOE-----KEDKPAETKKLRIANPP 534
DB 2228 TWIHIPGVNSSSSTSPVSKGPPLKTPASKSPSEGTATTPRGAKSVKSELSPVARQ 2287
QY 535 PTELGGSSGA-LEEGIKMSKWPPEDEISKPEVPEDVDLDLKLRLRSSSLKERS--RPF 591
DB 2288 TSQIGSSKAPSRSGSRDTPSRPAQOPLSRP-----IQSPGRNSISPRNGISPP 2338
QY 592 TAAASFQSTSVKSPKTVSPPIKAGMSMQSEESVGGVRAERKQVENAKSKNGNVGKT 651
DB 2339 NXLSQLPRTS--SPSTAS---TKSSGSGKVSYSFQRMQSQOQLTKQTGLSKNASSIPRS 2393
QY 652 TWQNKESKGTGKRSKEGHSLEMEENL---VENGADSDSDNSFLKQOS 698
DB 2394 ---ESASKGLNQMNNGANKKVELSRMSTKSSGSESDRSPVLRQS 2440
```

Db 2163 GPRILK-----PGEKSTLETXX----- 2179
QY 435 RIYCKPHNFQKSGNYDEGHRPHKDLWASK-NENEEI-----LERPAQ-----LANAR 485
Db 2180 -----IESESKGKIG-GKKVYKSLITGKVRNSSEISQMKQPLQANMPSISGR 2227
QY 486 ETPHSPGVED-----APIAKGVLAASMEAKASSQOE-----KEDKPAETKKLRIAMPP 534
Db 2228 TWIHIPGVNRSSSTSPVSKGPPKLTPTASKSPSEGQTATTSRGAKEPVSKSELSPVARQ 2287
QY 535 PTELGGSGSA-LEEGIKMVKPKWPPDEISKPEVPEDVDLKKLRSSSLKERS--RPF 591
Db 2288 TSQIGSGSKAPSRSGRSDTSPRPAQPLSRP-----ISPGNSISPGNIGISPP 2338
QY 592 TVAASFQSTSVKSPKTVPPPIRKGMNSEQSESVGGRVAERKQVENAKASKXGNVGT 651
Db 2339 NKLSOLPRTS--SPSTAS--TKSSGSKMSYTPSGRQMSQQLTKQTGLSKNASSIPRS 2393
QY 652 TWQNKESKGTGKRSKEGHSLEMEENL-----VENGADSDDDNSFLKQOS 698
Db 2394 ---ESASKGLNQMNNGANGANKVELSRMSSTKSSGESDRSERPVLVRQS 2440

RESULT 7

US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; US-08-452-654-7
Query Match 4.9%; Score 193; DB 1; Length 2842;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;
QY 26 VNNKSSAIVIFSKYQKAAEET--NMEKRSNTENLSQHFR----- 65
Db 1788 VRKNAD-----SKQNLNAERVFSDNKDKSKQNLKNNKDFNDKLPNNEDRVGRSFAFD 1840
QY 66 -----KGLTVLKKWENPGJCAESHTDSLANSSTEIRHADHPAE--VTSH---A 112
Db 1841 SPHHYTPTEGTPCFPSRNDLSLDDFDDDDVDLSREKAEELKAKENKSEAKVTSHTLT 1900
QY 113 ASGAXAQEQIHPRSLRSPPEALVQGRYPHIKDGEDLKH--STESKKVEN----- 163
Db 1901 SNQGSANKTQAIKQPINRGQPKPILOKQSTFPQSSKDIPIRGAATDEKLQNFATENTPV 1960
QY 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLMMFEKEGPTQTKILRAQSASGRK 223
Db 1961 CFSHNSLSLSLDDQENN-----NKENERIKETEPDSDQGEPSKPQ-----ASG-- 2005
QY 224 ISENSYSLDDLEI---GPGQLSSSTTFDEKNESSRNLELPRLSETSIKDRMAKYQAAVSK 280
Db 2006 YAPKSFHVEDTPVCFPSRNSLSUSJDSDD-----LLOECISSAMPK-----KK 2050
QY 281 QSSSTNYTNEL---KASGG-----EIK-IHKMEQKENVPPGPEVCITHQEGEKISAN 328
Db 2051 KPSRLKGDNEKXSPNMGGLIGEDLTLDLKDQIRPDSEHGLSPDSE-----NFDWKAQEG 2106
QY 329 ENSL-----AVRSTPAEDDSRDS--QVKSEVQ--QVHPKPLSPDGRASSLSSES 374
Db 2107 ANSIVSSLHQAAAAACLQRQASDS--DSILSKSGISLGSPPH---LTPDQEEKPFTSNK 2162
QY 375 PPKAMKKFCAPARETCVCOQTVYPMERLLANQOVFHISCFRCSCYCNKLSLGTIYASLHG 434
Db 2163 GPRILK-----PGEKSTLETXX----- 2179
QY 435 RIYCKPHNFQKSGNYDEGHRPHKDLWASK-NENEEI-----LERPAQ-----LANAR 485
Db 2180 -----IESESKGKIG-GKKVYKSLITGKVRNSSEISQMKQPLQANMPSISGR 2227
QY 486 ETPHSPGVED-----APIAKGVLAASMEAKASSQOE-----KEDKPAETKKLRIAMPP 534
Db 2228 TWIHIPGVNRSSSTSPVSKGPPKLTPTASKSPSEGQTATTSRGAKEPVSKSELSPVARQ 2287
QY 535 PTELGGSGSA-LEEGIKMVKPKWPPDEISKPEVPEDVDLKKLRSSSLKERS--RPF 591
Db 2288 TSQIGSGSKAPSRSGRSDTSPRPAQPLSRP-----ISPGNRSISPGNIGISPP 2338
QY 592 TVAASFQSTSVKSPKTVPPPIRKGMNSEQSESVGGRVAERKQVENAKASKXGNVGT 651
Db 2339 NKLSOLPRTS--SPSTAS--TKSSGSKMSYTPSGRQMSQQLTKQTGLSKNASSIPRS 2393
QY 652 TWQNKESKGTGKRSKEGHSLEMEENL-----VENGADSDDDNSFLKQOS 698
Db 2394 ---ESASKGLNQMNNGANGANKVELSRMSSTKSSGESDRSERPVLVRQS 2440

RESULT 8

US-08-449-731-7
; Sequence 7, Application US/08449731
; Patent No. 6413727
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA

HEDGE, PHILIP J.
JOSLYN, GEOFF
KINZLER, KENNETH
MARKHAM, ALEXANDER F.
NAKAMURA, YUSUKE
THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,731
FILING DATE: 25-May-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,548
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CLONE: APC
IMMEDIATE SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-449-731-7
Query Match 4.9%; Score 193; DB 4; Length 2842;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;
26 VNKVSSAIVEIFSYKAAET---NNEKKSNNTENLSQHFR----- 65
1788 VKNAD-----SKNLAERVFSDKSKQNKQNKNSKDFNDKLPNNEDVRGSAFD 1840
66 -----KGLTVLKKXWENPGLGAESHTSLRNSSTELRHADHPAE--VTSH---A 112
1841 SPHTVPTGTYCFYSRNDLSLDFDDDDVDLSREKAEKAKENKESEAKVTSHTLT 1900
113 ASGAKADQEQIHRSLRSPPEALVQGRYPHIKDGEDLKDHP-STESKKMEN----- 163
1901 SNQSQANKTAQAKQPINRGQPKPILOQSTFPQSSKDIPIRGAATDEKLQNFALIENTPV 1960
164 CLGESRHEVEKEISENTDASKIEKYVNLRLKQMFKEGPEPTQTKILRAQSASGRK 223
1961 CFSHNSLSLSLSDIQENN-----NKENEPIKETEPDQGEPSKPQ-----ASG-- 2005
224 ISENSYSLDDLEI---GFCQLSSSTFDSEKNSRRNLPLRLSETSIKDRNAKYQAIVSK 280
2006 YAPKGFHVEDTPVCFSRNSSSSLSIDSEDD-----LLOECISSAMPK-----KK 2050

281 QSSSTNYTNEL---KASGG-----EIK-IHKMEQKENVPPGPEVCITHQGEKISAN 328
2051 KPSRLKGDNEKHSRPNMGILGEDTLIDLKIDQRPDSEHGLSPDSE-----NFDWKAIQEG 2106
329 ENSL-----AVRSTPAEDSDSDS--QVKEVQ--QPVHPKPLSPDSRASSJESS 374
2107 ANSIVSSLHQAAAAACLSRQASDS-DSILSLKSGISLGSPPH---LTPQEEKPFTSNK 2162
375 PPRAMKXFOAPARETCVCEOKTVYPMERLLANOQVPHISCFRCSCYCNKLSLGTIASLHG 434
2163 GPRILK-----PGEKSTLTETK----- 2179
435 RIYCKPHFNOLFYSKGNVDEGFGRPHKDIWASK-NENEIE---LERPAO-----LANAR 485
2180 -----IESEKGIKG-GKKVYKSLITKVRNSNISEIGMKQKQLOANMPSISRGR 2227
486 ETPHSPGVED-----APIAKVGVLAAAMEAKASSQOE-----KEDKPAETTKKLRIAWPP 534
2228 TMIHFGVRNRSSTSTSPVSKKGPLKTPASKSSEGOATATSPRGAKPSVKSELSPVARQ 2287
535 PTELGGSSGA-LEEGIKMVKPWPPEDEISKPEVPEVDLDLKKLRSSSLKERS--RPF 591
2288 TSQIGGSSKAPSRSGSRDSTPSRPAQOPLSRP-----IQSPGRNSISPGRNGISPP 2338
592 TVAASFQSTSVKSPKTVSPPIRKGMSESESESVGGRVAERKQVENAKASKKNGNVGKT 651
2339 NKLSQLPRTS--SPSTAS---TKSSGSGMSYSPGRQMSQQNLTKQTGLSKQASSIPRS 2393
652 TWQNKESGKETGKSGEGHGLEMENENL---VENGADSDDEDDNSFLKQOS 698
2394 ---ESASKGLNQMNNGANGANKVELSRMSSTKSGSESDRSERPVLVRQS 2440
RESULT 9
US-07-741-940-2
Sequence 2, Application US/07741940
Patent No. 5352775
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: NAKAMURA, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-741-940-2

Query Match 4.9%; Score 193; DB 1; Length 2843;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKKNSSAIVFISKYQAAEET--NMKKRSNTENLSQHFR----- 65
Db 1789 VRKNAD-----SKNNLNARVFSNKKQNLKNNKDFNDKLPNNEDRVGSGAFD 1841
QY 66 -----KGTLLVKKWNPGLGAESHTDSLNSSTEIRHRADHPAE--VTSH---A 112
Db 1842 SPHYTPIEGTPYCFSRNDSLSLDPDDDDVLSREKAEURLKAKENKESAKVTSHTLT 1901
QY 113 ASGAKADQEQIHRSLRSLRSPPEALVQGRYPHIKDGEDLDKH-STESKWMEN----- 163
Db 1902 SNOOSANKTOAIKQPINRGQPKILQKSTFPQSKDIPDRGAATDEKLQNFALIENTPV 1961
QY 164 CLGSRHEVEKSEISENTDASGKIEKYNVPLNRLKWMFEKGEPTQTILRAQSRASGRK 223
Db 1962 CFSHNSLSLSLSDIDQENN-----NKNEPIKTEPPDQSGEPEKQP-----ASG-- 2006
QY 224 ISENSYSDDLLEI---GPGQLSSSTFSEKNSRRNLELPRLSETSIKDRMAKYQAAVSK 280
Db 2007 YAPKSFHVEDTPVCFSRNDSLSLSDSEDD-----LLOECISSAMPK-----KK 2051
QY 281 QSSSTNYTNEL---KASGG-----EIK-IHKMEQKENVPGPRVCITHQEGEKISAN 328
Db 2052 KPSRLKGDNEKHSRNNWGGILGEDLTLDKDIQRPDSEHGLSPDSE-----NFDWKAIQEG 2107
QY 329 ENSL-----AVRSTPAEDSDRS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
Db 2108 ANSIVSSLHQAAAAACLSRQSSDS-DSILSLKSGISLGSPFH---LTPDQEEKPFTSNK 2163
QY 375 PPKAMKFFQAPARETCVECKTVPMERLLANQOVFHSICFRCSYCNKLSLQTYASLHG 434
Db 2164 GPRLK-----PGKSTLETKK----- 2180
QY 435 RIYKPHFNQLFKSGNYDEGFGHRPHKDLWASK-NENEEI---LERPAQ-----LANAR 485
Db 2181 -----IESESKGIGK-GKKVYKSLITGKVRNSNLSIQMKQPLQANWPSISRGR 2228
QY 486 ETHSPQVED-----APIAKVGVLAAAMEAKASQOE-----KEDKPAETKKLRIAWPP 534
Db 2229 TMIHIPGVNRSSSTSPVSKGPPKLTTPASKSPSEGGTATTSRPAKSPVSKSELSPVARQ 2288
QY 535 PTELGGSSGSA-LEEGIKMSKPKWPEDEISKPEVPEDVDLKLKLRSSSLKERS--RPF 591
Db 2289 TSQIGGSKAPSRSGSRDSTSPRAQOPLSRP-----IOSPGNRSISPGNRISPP 2339
QY 592 TVAASFQSTSVKPKTVPPIRKGWMSQESSESVGGRVAERQVENAKKNGNVGKT 651
Db 2340 NKLSQLPRTS--SPSTAS---TKSSGSKMSYTPSGRQMSQNLTKQTGLSKNASSIPRS 2394
QY 652 TWNKESKGTGRSKRGHGLENNENL-----VENGADSDDEDSFLKQOS 698
Db 2395 ---ESASKGLNMNGNGANKKVELRMSSTKSGSSESRSPVLVRQS 2441

RESULT 10
US-08-289-548A-2
Sequence 2, Application US/08289548A
Patent No. 5648212
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-289-548A-2

Query Match 4.9%; Score 193; DB 1; Length 2843;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKKNSSAIVFISKYQAAEET--NMKKRSNTENLSQHFR----- 65
Db 1789 VRKNAD-----SKNNLNARVFSNKKQNLKNNKDFNDKLPNNEDRVGSGAFD 1841
QY 66 -----KGTLLVKKWNPGLGAESHTDSLNSSTEIRHRADHPAE--VTSH---A 112
Db 1842 SPHYTPIEGTPYCFSRNDSLSLDPDDDDVLSREKAEURLKAKENKESAKVTSHTLT 1901
QY 113 ASGAKADQEQIHRSLRSLRSPPEALVQGRYPHIKDGEDLDKH-STESKWMEN----- 163
Db 1902 SNOOSANKTOAIKQPINRGQPKILQKSTFPQSKDIPDRGAATDEKLQNFALIENTPV 1961
QY 164 CLGSRHEVEKSEISENTDASGKIEKYNVPLNRLKWMFEKGEPTQTILRAQSRASGRK 223
Db 1962 CFSHNSLSLSLSDIDQENN-----NKNEPIKTEPPDQSGEPEKQP-----ASG-- 2006
QY 224 ISENSYSDDLLEI---GPGQLSSSTFSEKNSRRNLELPRLSETSIKDRMAKYQAAVSK 280
Db 2007 YAPKSFHVEDTPVCFSRNDSLSLSDSEDD-----LLOECISSAMPK-----KK 2051
QY 281 QSSSTNYTNEL---KASGG-----EIK-IHKMEQKENVPGPRVCITHQEGEKISAN 328
Db 2052 KPSRLKGDNEKHSRNNWGGILGEDLTLDKDIQRPDSEHGLSPDSE-----NFDWKAIQEG 2107
QY 329 ENSL-----AVRSTPAEDSDRS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
Db 2108 ANSIVSSLHQAAAAACLSRQSSDS-DSILSLKSGISLGSPFH---LTPDQEEKPFTSNK 2163

```
QY 375 PPKAMKFFQAPARETCVECKTYPYMERLLANQOVPHISCFRCSYCNKLSLGTYSALHG 434
Db 2164 GPRILK-----PGEKSTLETK-----2180
QY 435 RIYCKPHNFQKSKNGYDEGFHGRPHKDLWASK-NENEEI---LERPAQ-----LANAR 485
Db 2181 -----ISESKGIGK-GKKVYKSLITGKVRNSSEISQGMKQPLQANMPSISRGR 2228
QY 486 ETPHSGVED-----APIAKGVGLAASMEAKASSQOE-----KEDKPAETKKLRIAMPP 534
Db 2229 TMIHIGVRNSSSTSPVSKGKPLTPASKSPSEGOATTTPRGAKPSVKSELSPVARQ 2288
QY 535 PTELGGSSGA-LLEGIKMSPKPPPEDEISKPEVPEDVDLDLKKLRSSSLKERS--RPF 591
Db 2289 TSQIGSSKAPSRSGSRDSTPSRPAQPLSRP-----IQSPGRNSISPGRNGISPP 2339
QY 592 TVAAASFQSTSVKSPKTVSPPIRKGMSESESVGGRVAERKQVENAKSKNGNVGKT 651
Db 2340 NKLSQLPRTS--SPSTAS---TKSSSGKMSYTPSGRQMSQQLTKQTGLSKNASSIPRS 2394
QY 652 TWQKESKGETGKRSKEGHSLENNENL---VENGADSDDDNSFLKQOS 698
Db 2395 ---ESASKGLNQMNNGNGANKKVELSRMSSTKSSGESDRSERPVLVRQS 2441

RESULT 11
US-08-452-654-2
; Sequence 2, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
```

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-452-654-2

Query Match 4.9%; Score 193; DB 1; Length 2843;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 293; Indels 214; Gaps 37;

QY 26 VNKKSSAIVEIFSKYKAAEET---NMEKKRSNTENLSQHFR-----65
Db 1789 VRKNAD-----SKNNLNAERVSFNDKDSKKQNLKNNKDFNDKLPNNEDRVGRSPAFD 1841
QY 56 -----KGTLTVLVKKWENPGLGAESHDTLSLNSSTEIRHADHPAE--VTSH---A 112
Db 1842 SPHYTYPIETGTPCYFSRNDLSLSLQFDDDDVDLSREKAEIRKAKENKESAEKVTSHTELT 1901
QY 113 ASGAKADEEIOIHPRSLRSLRSPREALVOGRYPHIKDGDLDKDH-STESKSMEN-----163
Db 1902 SNOQSANKTOIAIKQPINRGQPKFILOKQSTFPOSSKDIPDRGAATDEKLNFAIENTPV 1961
QY 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTQKILRAQSASGRK 223
Db 1962 CFSHNSLSLSLSDIDQENN-----NKENEPKETETPPDSQGEPSKPQ-----ASG-- 2006
QY 224 ISENSYSLDDLEI---GPGQLSSSTFDSERKNESRNLLELPLRLSETSIKDRMAKYQAAVSK 280
Db 2007 YAPKSFHVEDTPVCFSRNSSLSSLSIDSEDD-----LLQECISSAMPK-----KK 2051
QY 281 QSSSTNYTNEI---KASGG-----EIK-IHKMEOKENVPPGPEVCITHQEGEKISAN 328
Db 2052 KPSRLKGDNEKHSRPNMGIGLEDTLTLDKDIQRPDSEHGLSPDSE---NFDWKATQEG 2107
QY 329 ENSL-----AVRSTPAEDDSRDS--QVKEVQ--QPVHPKPLSPDSRASSLSSESS 374
Db 2108 ANSIVSSLHQAAAACLSRQASSDS-DSILSLKSGISLGSPPH---LTPDOEEKPFTSNK 2163
QY 375 PPKAMKFFQAPARETCVECKTYPYMERLLANQOVPHISCFRCSYCNKLSLGTYSALHG 434
Db 2164 GPRILK-----PGEKSTLETK-----2180
QY 435 RIYCKPHNFQKSKNGYDEGFHGRPHKDLWASK-NENEEI---LERPAQ-----LANAR 485
Db 2181 -----ISESKGIGK-GKKVYKSLITGKVRNSSEISQGMKQPLQANMPSISRGR 2228
QY 486 ETPHSGVED-----APIAKGVGLAASMEAKASSQOE-----KEDKPAETKKLRIAMPP 534
Db 2229 TMIHIGVRNSSSTSPVSKGKPLTPASKSPSEGOATTTPRGAKPSVKSELSPVARQ 2288
QY 535 PTELGGSSGA-LLEGIKMSPKPPPEDEISKPEVPEDVDLDLKKLRSSSLKERS--RPF 591
Db 2289 TSQIGSSKAPSRSGSRDSTPSRPAQPLSRP-----IQSPGRNSISPGRNGISPP 2339
QY 592 TVAAASFQSTSVKSPKTVSPPIRKGMSESESVGGRVAERKQVENAKSKNGNVGKT 651
Db 2340 NKLSQLPRTS--SPSTAS---TKSSSGKMSYTPSGRQMSQQLTKQTGLSKNASSIPRS 2394
QY 652 TWQKESKGETGKRSKEGHSLENNENL---VENGADSDDDNSFLKQOS 698
Db 2395 ---ESASKGLNQMNNGNGANKKVELSRMSSTKSSGESDRSERPVLVRQS 2441

RESULT 12
US-08-452-655B-2
; Sequence 2, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
```


TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
US-08-452-655B-7

Query Match 4.9%; Score 193; DB 1; Length 2843;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKXSSAIVEIFSKYKAAEET---NMKKRSNTENLSQHFR----- 65
Db 1789 VRKNAD-----SKNUNLAERVFSKQKQNLKNNKSKDFNDKLPNNEDVRGSAFDP 1841
QY 66 -----KGTITLVKKWENPGLGAESHTDLSRNSSTEIRHRAHPAE--VTSH---A 112
Db 1842 SPHHYPIEGTPTCFSRNDSLSLDFDDDDVLSREKAELEKAKENKESAKVTSHTELT 1901
QY 113 ASGAKADQEQIHPHRSRLSPPEALVQGRYPHIKDGEDLKH-STESKKNEN----- 163
Db 1902 SQQQANKTQAIKQPIRNGQPKILQKSTFPQSSKDIPDGAATDEKQNFATENTPV 1961
QY 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRASGRK 223
Db 1962 CFSHNSLSLSLDIQENN-----NKNEPIKETETPPDSQGEPSKPQ-----ASG-- 2006
QY 224 ISENSYSLDDLEI---GPGQLSSSTFDEKNSRRNLELPRSETSIKDRMAKYQAASVK 280
Db 3007 YAPKSFHVEDTPVCFSRNDSLSLSDSDD-----LQECISSAMPK-----KK 2051
QY 281 QSSSTNYTNEL---KASG-----EIK-IHKMEQKENVPPGPEVCITHOEGEKISAN 328
Db 2052 KPSRLKGDNEKHSRNMGGILGEDLTLDKDIQRPDSEHGLSPDSE-----NFDWKAIQEG 2107
QY 329 ENSL-----AVRSTPAEDDSRDS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
Db 2180 ANSIVSSLHQAAAAACLSRQASDS--DSILSKSGISLSPFH---LTPDQEEKPFTSNK 2163
QY 375 PPKAMKKFQAPARETCVQCQKTPVPMERLLANQVPHISCPRCSYCNKNKLSLGTVASLHG 434
Db 2164 GPRILK---PQEKSTLETK----- 2180
QY 435 RIYKPHFNQLEKSKGNVDGEGHPRPHKDLWASK-NENEEL---LERPAQ-----LANAR 485
Db 2181 -----IESESKGIKG-GKKVYKSLITGKVRNSNISEISGQMKQPLQANMPSISGR 2228
QY 486 ETPHSPGVED-----APIAKGVLAASMEAKASSQOE-----KEDKPAETKRLIAWPP 534
Db 2229 TWIHIPGVNSSSTSPVSKGPPKTPASKSPSEGQTATSPRGAKPSVKSLSPLVARQ 2288
QY 535 PTELSSGSA-LEEGIKMSKPKWPPEDISKEPVEDVDLKLRRSSSLKERS--RPF 591
Db 2289 TSQIGSSKAPSRGSRDSTSPRPAQPLSRP-----IQSPGRNSISPGRANGISPP 2339
QY 592 TVAASFQSTSVKSPKTPVSPTRKGNMSGEQEEESVGGRAVERKQVENAKSKKXGNVGKT 651
Db 2340 NKLSQLPRTS--SPETAS---TKSSGSGKMSYTPSGRQMSQONLTKQTGLSKNASSIPRS 2394
QY 652 TWQNKESKETGRSKEGHSLEMEENEL---VENGADSDDEDNLFKQOS 698
Db 2395 ---ESAGKGLQNMNGNGANKKVELLSMSTSKSGSGESDRSERPVLVRQS 2441

RESULT 14

US-08-370-235A-2

; Sequence 2, Application US/08370235A

Patent No. 5910418
GENERAL INFORMATION:
APPLICANT: VOGELSTEIN, BERT
APPLICANT: KINZLER, KENNETH W.
APPLICANT: HILL, DAVID E.
APPLICANT: JOHNSON, KAREN A.
TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DETERMINING
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: BANNER & WITCOFF, LTD.
STREET: 1001 G STREET, N.W.
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,235A
FILING DATE: 01-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KAGAN, SARAH A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107.48688
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 508 9100
TELEFAX: 202 508 9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-370-235A-2

Query Match 4.9%; Score 193; DB 2; Length 2843;

Best Local Similarity 19.4%; Pred. No. 2e-06;

Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

QY 26 VNKXSSAIVEIFSKYKAAEET---NMKKRSNTENLSQHFR----- 65
Db 1789 VRKNAD-----SKNUNLAERVFSKQKQNLKNNKSKDFNDKLPNNEDVRGSAFDP 1841
QY 66 -----KGTITLVKKWENPGLGAESHTDLSRNSSTEIRHRAHPAE--VTSH---A 112
Db 1842 SPHHYPIEGTPTCFSRNDSLSLDFDDDDVLSREKAELEKAKENKESAKVTSHTELT 1901
QY 113 ASGAKADQEQIHPHRSRLSPPEALVQGRYPHIKDGEDLKH-STESKKNEN----- 163
Db 1902 SQQQANKTQAIKQPIRNGQPKILQKSTFPQSSKDIPDGAATDEKQNFATENTPV 1961
QY 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSRASGRK 223
Db 1962 CFSHNSLSLSLDIQENN-----NKNEPIKETETPPDSQGEPSKPQ-----ASG-- 2006
QY 224 ISENSYSLDDLEI---GPGQLSSSTFDEKNSRRNLELPRSETSIKDRMAKYQAASVK 280
Db 2007 YAPKSFHVEDTPVCFSRNDSLSLSDSDD-----LQECISSAMPK-----KK 2051
QY 281 QSSSTNYTNEL---KASG-----EIK-IHKMEQKENVPPGPEVCITHOEGEKISAN 328
Db 2052 KPSRLKGDNEKHSRNMGGILGEDLTLDKDIQRPDSEHGLSPDSE-----NFDWKAIQEG 2107
QY 329 ENSL-----AVRSTPAEDDSRDS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
Db 2108 ANSIVSSLHQAAAAACLSRQASDS--DSILSKSGISLSPFH---LTPDQEEKPFTSNK 2163
QY 375 PPKAMKKFQAPARETCVQCQKTPVPMERLLANQVPHISCPRCSYCNKNKLSLGTVASLHG 434

Db 2164 GPRILK-----PGKSTLETK-----2180
Qy 435 RIYCKPHNFQKSGNYDEGFGHRRPHKDLWASK-NENEEI-----LERPAQ-----LANAR 485
Db 2181 -----IESESKGIG-CKVKYKSLITGKVRNSENSEISQGMKQPLQANMPSISGR 2228
Qy 486 ETHSPGVED-----APIAKGVGLAASMEAKASSQOE-----KEDKPAETKKLRIAMPP 534
Db 2229 TMIHIPGVNRSSSTSPVSKGKPLKTPASKSPSEGGTATTSPRGAKPSVKSELSPVARQ 2288
Qy 535 PTELGGSSGA-LEEGIKMSPKWPPEDEISKPEVPEDVDLDLKKLRSSSLKERS--RPF 591
Db 2289 TSQIGSSKAPSRSGRSDTSPSPAQPLSRP-----IOSPGNISISPGNISIPP 2339
Qy 592 TVAASFQSTSVKSPKTVPPPIRGKWSNSEQSEESVGGVRAERKQVENAKASKNGNVGKT 651
Db 2340 NKLSQLPRTS--SPSTAS---TKSSGGMKSYTSPGRQMSQQLTKQTGLSKNASSIPRS 2394
Qy 652 TWQNKESKGTGKRSKEGHSLEMEENL-----VENGADSDDDNSFLKQOS 698
Db 2395 ---ESASKGLNQMNNGNGANKKVELSRMSSTKSGSSESDRSPVLRQS 2441

RESULT 15

US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GREGG
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MAKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-450-582-2

Query Match 4.9%; Score 193; DB 3; Length 2843;
Best Local Similarity 19.4%; Pred. No. 2e-06;
Matches 149; Conservative 124; Mismatches 283; Indels 214; Gaps 37;

Qy 26 VNKKSSAIVFISKYQAAEET---NMEKRSNTENLSQHFR-----65
Db 1789 VRKNAD-----SKNVLNAERFSDNKKSKQKLNKSKDFNDKLPNNEDVRGSAFD 1841
Qy 66 -----KGTLLVLKKWENPGLGAESHTDSLNSSTEIHRADHPAE--VTSH---A 112
Db 1842 SPHYPTPIEGTPYCFSRNDSLSLDFDODDDVDLSREKAELKAKENKESAKVTSHTLT 1901
Qy 113 ASGAKADQEEQIHPRSLRSPPEALVQGRYPHIKDGEDLDKH-STESKKMEN-----163
Db 1902 SNQGSANKTQAIKQPINRGQPKILOKQSTFPOSSKDIIDRGAATDEKLQNFALIENTPV 1961
Qy 164 CLGESRHEVEKSEISENTDASGKIEKYNVPLNRLKMMFEKGEPTQTKILRAQSASGRK 223
Db 1962 CFSHNSLSLSLSDIDQENN---NKENEPIKETEPDPOSQGEPSKPQ-----ASG--2006
Qy 224 ISENSYSLDDLEI---GPCQLSSSTFSEKESRNLELPLSETSIKDRMAKQAASVK 280
Db 2007 YAPKSFHVEDTFVCFSRNSSLSDISEDD-----LLOECISSAMPK-----KK 2051
Qy 281 OSSSTNYTNEL---KASGG-----EIK-IHKVEOKENVPPGPEVCITHOEGEKISAN 328
Db 2052 KPSRLKGDNEKHSRNMGGLGEBLTLDLKIQPDSEHGLSPDSE-----NFDKAIQEG 2107
Qy 329 ENSL-----AVRSTPAEDDSRDS--QVKSEVQ--QPVHPKPLSPDSRASSLSSESS 374
Db 2108 ANSIVSSLHQAAAAACLSRQASSDS--DSILSKSGISLGSPPH---LTPDQEEKPFTSNK 2163
Qy 375 PPKAMKFOAPARETCVECOQTVPMERLLANQOVFHIISCPRCYCNNKLSLGTYSALHG 434
Db 2164 GPRILK-----PGKSTLETK-----2180
Qy 435 RIYCKPHNFQKSGNYDEGFGHRRPHKDLWASK-NENEEI---LERPAQ-----LANAR 485
Db 2181 -----IESESKGIG-CKVKYKSLITGKVRNSENSEISQGMKQPLQANMPSISGR 2228
Qy 486 ETHSPGVED-----APIAKGVGLAASMEAKASSQOE-----KEDKPAETKKLRIAMPP 534
Db 2229 TMIHIPGVNRSSSTSPVSKGKPLKTPASKSPSEGGTATTSPRGAKPSVKSELSPVARQ 2288
Qy 535 PTELGGSSGA-LEEGIKMSPKWPPEDEISKPEVPEDVDLDLKKLRSSSLKERS--RPF 591
Db 2289 TSQIGSSKAPSRSGRSDTSPSPAQPLSRP-----IOSPGNISISPGNISIPP 2339
Qy 592 TVAASFQSTSVKSPKTVPPPIRGKWSNSEQSEESVGGVRAERKQVENAKASKNGNVGKT 651
Db 2340 NKLSQLPRTS--SPSTAS---TKSSGGMKSYTSPGRQMSQQLTKQTGLSKNASSIPRS 2394
Qy 652 TWQNKESKGTGKRSKEGHSLEMEENL-----VENGADSDDDNSFLKQOS 698
Db 2395 ---ESASKGLNQMNNGNGANKKVELSRMSSTKSGSSESDRSPVLRQS 2441

Search completed: January 6, 2004, 09:48:58
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:47:24 ; Search time 38 Seconds

(without alignments)
4007.544 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSFPNRRQWTSLSLRVTA.....LSVEQIKENRYDEDEEE 759

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3869	98.5	755	11	US-09-783-732-4
2	3027.5	77.1	596	11	US-09-783-732-2
3	1254	31.9	243	9	US-09-925-297-611
4	464.5	11.8	527	9	US-09-880-192-53
5	464.5	11.8	527	12	US-10-427-348-53
6	363	9.2	158	9	US-09-789-919-42
7	357	9.1	127	12	US-10-440-366-12
8	357	9.1	127	12	US-10-144-047-2959
9	281	7.2	52	11	US-09-783-732-5
10	245	6.2	1965	12	US-10-359-012-6
11	223	5.7	1312	12	US-10-126-704-29
12	223	5.7	1312	15	US-10-071-179-29
13	222	5.7	2476	11	US-09-824-574-7
14	221	5.6	4723	12	US-10-359-012-8
15	213	5.4	1633	12	US-10-359-012-4

Sequence 5013, Ap
Sequence 22, Appl
Sequence 1586, Ap
Sequence 103, Appl
Sequence 84, Appl
Sequence 16, Appl
Sequence 105, Appl
Sequence 10, Appl
Sequence 20, Appl
Sequence 6328, Ap
Sequence 30, Appl
Sequence 7, Appl
Sequence 8305, Ap
Sequence 62, Appl
Sequence 104, Appl
Sequence 34248, A
Sequence 423, Appl
Sequence 116, Appl
Sequence 6, Appl
Sequence 32, Appl
Sequence 30, Appl
Sequence 30, Appl
Sequence 17, Appl
Sequence 1, Appl
Sequence 16, Appl
Sequence 3, Appl
Sequence 13, Appl
Sequence 7646, Ap

ALIGNMENTS

RESULT 1

US-09-783-732-4
; Sequence 4, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; TITLE OF INVENTION: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; TITLE OF INVENTION: (EPLIN)
; FILE REFERENCE: 10809/03001
; CURRENT APPLICATION NUMBER: US/09783,732
; PRIOR FILING DATE: 2001-02-13
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homosapien
US-09-783-732-4

Query Match 98.5%; Score 3869; DB 11; Length 755;
Best Local Similarity 99.5%; Pred. No. 1.6e-250;
Matches 755; Conservative 0; Mismatches 0; Indels 4; Gaps 4;

Qy	1	MESSFPNRRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMKRSNTENL	60
Db	1	MESSFPNRRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYQKAAEETNMKRSNTENL	60
Qy	61	SQFRKGLTVLKKKENPGLGAESHTDSLRNSTIRHRAHPHPPAVTSHAASGAKAQ	120
Db	61	SQFRKGLTVLKKKENPGLGAESHTDSLRNSTIRHRAHPHPPAVTSHAASGAKAQ	120
Qy	121	EQIHPRSLRSPPEALVQGRYPHIKDGELDKDHSTESKKNENCLGSRHEVEKSIEN	180
Db	121	EQIHPRSLRSPPEALVQGRYPHIKDGELDKDHSTESKKNENCLGSRHEVEKSIEN	180

191 TDAGKIEKYNVPLNRLKMFKEGPTQTQKILRAQSRASGRKISSENSYSLDLEIGPGQ 240
181 TDAGKIEKYNVPLNRLKMFKEGPTQTQKILRAQSRASGRKISSENSYSLDLEIGPGQ 240
241 LSSSTFSEKESRRNLPLSETSIKORMAKYQAAVSKQSSSTNTNKLKASGGI 300
241 LSSSTFSEKESRRNLPLSETSIKORMAKYQAAVSKQSSSTNTNKLKASGGI 300
301 HKMEQKENVPPGPEVCITHQGEKISANENSIAVRSSTPAEDSDSDSVKSEVQPVHPKP 360
301 HKMEQKENVPPGPEVCITHQGEKISANENSIAVRSSTPAEDSDSDSVKSEVQPVHPKP 360
361 LSPDSRASSLSSESPPKAMKFFQAPARETCVECKTVPYMERLLANQQVFIHSCFRCSYC 420
361 LSPDSRASSLSSESPPKAMKFFQAPARETCVECKTVPYMERLLANQQVFIHSCFRCSYC 420
421 NNKLSGTGYASLHGRIYCKPHFNQLFKSKGNYDEGFGHRRPKDLWASKNEEILERPAP 480
421 NNKLSGTGYASLHGRIYCKPHFNQLFKSKGNYDEGFGHRRPKDLWASKNEEILERPAP 480
481 LANARETPHSPGVEDAPIAKVGLAASWEAKASSQOEKEDKPAETKKLRIAWPPPTLGS 540
481 LANARETPHSPGVEDAPIAKVGLAASWEAKASSQOEKEDKPAETKKLRIAWPPPTLGS 540
541 SGSALEEGIKMSKPKWPPDEISKPEVPEDVDLKLKRRSSSLKRSRPTTVAASFQST 600
541 SGSALEEGIKMSKPKWPPDEISKPEVPEDVDLKLKRRSSSLKRSRPTTVAASFQST 600
601 SVKSPKTVSPPIRGWMSQSESVGGRVAERQVENAKSKNGNVGKTTWQNKESKG 660
601 SVKSPKTVSPPIRGWMSQSESVGGRVAERQVENAKSKNGNVGKTTWQNKESKG 660
661 ETGRKSKEGHSLMENENLVENGADSDSDNSFLKQSPQOEPRKLNWSSFDVNTFAEEFT 720
660 ETGRKSKEGHSLMENENLVENGADSDSDNSFLKQ-PQEPK-LNWSSFDVNTFAEEFT 716
721 TQNKQSDQVELWEVVKELSVESQIKRNRYYDEDEDEE 759
717 TQNKQSDQVELWEVVKELSVESQIKRNRYYDEDEDEE 755

RESULT 2
US-09-783-732-2
; Sequence 2, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09/783,732
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homosapien
US-09-783-732-2

Query Match 77.1%; Score 3027.5; DB 11; Length 596;
Best Local Similarity 99.0%; Pred. No. 2.4e-194;
Matches 594; Conservative 0; Mismatches 1; Indels 5; Gaps 5;
161 MENCGLGSRHEVEKSEISENTDASGKIEKYNVPLNRLKMFKEGPTQTQKILRAQSRAS 220
1 MENCGLGSRHEVEKSEISENTDASGKIEKYNVPLNRLKMFKEGPTQTQKILRAQSRAS 60

221 GRKISENSYSLDLEIGPQQLSSSTFDSEKNSRRNLPLRLSETSIKORMAKYQAAVSK 280
61 GRKISENSYSLDLEIGPQQLSSSTFDSEKNSRRNLPLRLSETSIKORMAKYQAAVSK 120
281 QSSSTNTYTNELKASGGEIKIHMEQKENVPPGPEVCITHQGEKISANENSIAVRSSTPAE 340
121 QSSSTNTYTNELKASGGEIKIHMEQKENVPPGPEVCITHQGEKISANENSIAVRSSTPAE 180
341 DDS-RDSQVKSQVQPVHPKPLSPDSRASSLSSESPPKAMKFFQAPARETCVECKTVPY 399
181 DSDPGDSQVKSQVQPVHPKPLSPDSRASSLSSESPPKAMKFFQAPARETCVECKTVPY 240
400 MERLLANQQVFIHSCFRCSYCNKLSLGTGYASLHGRIYCKPHFNQLFKSKGNYDEGFGHR 459
241 MERLLANQQVFIHSCFRCSYCNKLSLGTGYASLHGRIYCKPHFNQLFKSKGNYDEGFGHR 300
460 PHKDLWASKNEEILERPAPLANARETPHSPGVEDAPIAKVGLAASWEAKASSQOEKE 519
301 PHKDLWASKNEEILERPAPLANARETPHSPGVEDAPIAKVGLAASWEAKASSQOEKE 360
520 DXPAETKKLRIAWPPPTLGGSSGSALEEGIKMSKPKWPPDEISKPEVPEDVDLKLKLR 579
361 DXPAETKKLRIAWPPPTLGGSSGSALEEGIKMSKPKWPPDEISKPEVPEDVDLKLKLR 420
580 RSSSLKRSRPTTVAASFQSTSVKSPKTVSPPIRGWMSQSESVGGRVAERQVENA 639
421 RSSSLKRSRPTTVAASFQSTSVKSPKTVSPPIRGWMSQSESVGGRVAERQVENA 480
640 KASKNGNVGKTTWQNKESKGKGTGRKSKEGHSLMENENLVENGADSDSDNSFLKQSP 699
481 KASKNGNVGKTTWQNKESKGKGTGRKSKEGHSLMENENLVENGADSDSDNSFLKQ-P 537
700 QEPKSLNWSSFDVNTFAEEFTTQNKQSDQVELWEVVKELSVESQIKRNRYYDEDEDEE 759
538 QEPK-LNWSSFDVNTFAEEFTTQNKQSDQVELWEVVKELSVESQIKRNRYYDEDEDEE 596

RESULT 3
US-09-925-297-611
; Sequence 611, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 611
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (185)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (237)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (238)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (243)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-611

Query Match 31.9%; Score 1254; DB 9; Length 243;

Best Local Similarity 98.8%; Pred. No. 3.5e-76;
Matches 238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 321 EGEKISANENSLAVRSTPAEDDSRDQVXSEVQVPHKPLSPDSRASSLSSESSPPKAMK 380
DB 1 EGEKISANENSLAVRSTPAEDDSRDQVXSEVQVPHKPLSPDSRASSLSSESSPPKAMK 60
QY 381 KFOAPARETCVECKTVYMERLLANQOVPHISCFRCSCYCNKLSLGTVYASLHGRYCKP 440
DB 61 KFOAPARETCVECKTVYMERLLANQOVPHISCFRCSCYCNKLSLGTVYASLHGRYCKP 120
QY 441 HFNLFKSKGNYDEGFGHPRKDLWASKNEEBILERPQALANARETPHSPGVEDAPIAK 500
DB 121 HFNLFKSKGNYDEGFGHPRKDLWASKNEEBILERPQALANARETPHSPGVEDAPIAK 180
QY 501 VGVLAASMEAKASQOEKEDKPAETKKLRIANPPPTLSSGSALEBEGIKMSKPKWPPED 560
DB 181 VGVLAASMEAKASQOEKEDKPAETKKLRIANPPPTLSSGSALEBEGIKMSKPKWXXED 240
QY 561 E 561
DB 241 E 241

RESULT 4
US-09-880-192-53
; Sequence 53, Application US/09880192
; Patent No. US20020077470A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/09/880,192
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 53
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020077470A1 058201CD1
US-09-880-192-53

Query Match 11.8%; Score 464.5; DB 9; Length 527;
Best Local Similarity 35.7%; Pred. No. 6.9e-23;
Matches 107; Conservative 47; Mismatches 105; Indels 41; Gaps 8;
QY 400 MERLLANQOVPHISCFRCSCYCNKLSLGTVYASLHGRYCKPHFNOLFKSKGNYDEGFGHR 459
DB 1 MECLVADKQNFHKSFCFRCHCNKSLNYSALHGOIYCKPHFKQLFKSKGNYDEGFGHK 60
QY 460 PHKDLWASKNEEBI-----LERPAQLANARETPHSPGVEDAPIAKVGVLAASMEAKASQ 515
DB 61 QHQRWNCNQSRSVDPINEEPNCKNTAENTLVPGDRNEHL-----DAGNSE 109
QY 516 QEKED--KPAETKKLRIANPPPTLSSGSALEBEGIKMSKPKWPE-DEISKPEVPEDVD 572
DB 110 GORNDRKLGERGKLVWPPSKPIKTLPPFEBELKMSKPKWPEMTLLSPFKSESL 169
QY 573 LDLLKLRSSSLKERSRPTVAASFQSTSVKSPKTVSPPIRKGMSEMSQSESVGGRVAE 632
DB 170 LEDVTPENKQRODHFFP-LQPYLQSTHVC-----QKEDVIG--IKE 209
QY 633 RKQVENAKSKNGNKGKTTWQNKESKGETKRSKEGHSLEMENENLVENGADSDDDNS 692
DB 210 MKMPEGRKDEKKE---GRKNVQDRPSEADTKSNRKSAMDNDNNVIVQSAEKEKNEKT 266

RESULT 6
US-09-789-919-42
; Sequence 42, Application US/09789919
; Patent No. US20020064855A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Kateri
; APPLICANT: Lemischka, Ihor
; TITLE OF INVENTION: GENES THAT REGULATE HEMATOPOIETIC BLOOD FORMING STEM
; FILE REFERENCE: 2275-1-005
; CURRENT APPLICATION NUMBER: US/09/789,919
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 42
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-789-919-42

RESULT 5
US-10-427-348-53
; Sequence 53, Application US/10427348
; Publication No. US20030175795A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-2 CON
; CURRENT APPLICATION NUMBER: US/10/427,348
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/880,192
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: US 09/299,708
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 53
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175795A1 058201CD1
US-10-427-348-53

Query Match 11.8%; Score 464.5; DB 12; Length 527;
Best Local Similarity 35.7%; Pred. No. 6.9e-23;
Matches 107; Conservative 47; Mismatches 105; Indels 41; Gaps 8;
QY 400 MERLLANQOVPHISCFRCSCYCNKLSLGTVYASLHGRYCKPHFNOLFKSKGNYDEGFGHR 459
DB 1 MECLVADKQNFHKSFCFRCHCNKSLNYSALHGOIYCKPHFKQLFKSKGNYDEGFGHK 60
QY 460 PHKDLWASKNEEBI-----LERPAQLANARETPHSPGVEDAPIAKVGVLAASMEAKASQ 515
DB 61 QHQRWNCNQSRSVDPINEEPNCKNTAENTLVPGDRNEHL-----DAGNSE 109
QY 516 QEKED--KPAETKKLRIANPPPTLSSGSALEBEGIKMSKPKWPE-DEISKPEVPEDVD 572
DB 110 GORNDRKLGERGKLVWPPSKPIKTLPPFEBELKMSKPKWPEMTLLSPFKSESL 169
QY 573 LDLLKLRSSSLKERSRPTVAASFQSTSVKSPKTVSPPIRKGMSEMSQSESVGGRVAE 632
DB 170 LEDVTPENKQRODHFFP-LQPYLQSTHVC-----QKEDVIG--IKE 209
QY 633 RKQVENAKSKNGNKGKTTWQNKESKGETKRSKEGHSLEMENENLVENGADSDDDNS 692
DB 210 MKMPEGRKDEKKE---GRKNVQDRPSEADTKSNRKSAMDNDNNVIVQSAEKEKNEKT 266

Query Match 9.2%; Score 363; DB 9; Length 158;
Best Local Similarity 58.7%; Pred. No. 8.4e-17;
Matches 64; Conservative 16; Mismatches 27; Indels 2; Gaps 1;

QY 362 SPDSRASSLSPPKAMKF--QAPARETCVQKTYPMERLLANQVPHISCFRCV 419
Db 11 TPSHEAKSGSGSTVQSKSFSLRAQVKETCAACQKTYPMERLVADKLIFHNSCFCKH 70

QY 420 CNKLSLCTYASLGRYCKPHFNQLFKSKGNVDEGFGHRPHKDLWASK 468
Db 71 CHTKLSLGSYAAMHGEFYCRPHFQQLFKSKGNVDEGFGRKQHKELWAHK 119

RESULT 7
US-10-440-366-12
; Sequence 12, Application US/10440366
; Publication No. US20030181706A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Voikmuth, Wayne
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: INFLAMMATION-ASSOCIATED POLYNUCLEOTIDES
; FILE REFERENCE: PB-0006-1 CIP
; CURRENT APPLICATION NUMBER: US/10/440,366
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: US 09/855,323
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/195,292
; PRIOR FILING DATE: 1998-11-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PERL Program
; SEQ ID NO 12
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 402234CD1
US-10-440-366-12

Query Match 9.1%; Score 357; DB 12; Length 127;
Best Local Similarity 50.7%; Pred. No. 1.6e-16;
Matches 69; Conservative 16; Mismatches 27; Indels 24; Gaps 3;

QY 333 AVRSTPAEDDRSDQVSEVQVPHPKPLSPDSRASSLSSESSPPKAMKXQAPARETCVE 392
Db 7 AAQATPSHDAKGGG--SSTVQ-----RSKSPS-----LRAQVKETCAA 42

QY 393 CQKTYPMERLLANQVPHISCFRCVYCNKLSLGTYSASLHGRYCKPHFNQLFKSKGNV 452
Db 43 CQKTYPMERLVADKLIFHNSCFCKCHTKLSLGSYAALHGEFYCKPHFQQLFKSKGNV 102

QY 453 DEGFGRPHKDLWASK 468
Db 103 DEGFGRKQHKELWAHK 118

RESULT 8
US-10-104-047-2959
; Sequence 2959, Application US/10104047
; Publication No. US200303392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US200303392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2959
; LENGTH: 127
; TYPE: PRT

; ORGANISM: Homo sapiens
US-10-104-047-2959

Query Match 9.1%; Score 357; DB 12; Length 127;
Best Local Similarity 50.7%; Pred. No. 1.6e-16;
Matches 69; Conservative 16; Mismatches 27; Indels 24; Gaps 3;

QY 333 AVRSTPAEDDRSDQVSEVQVPHPKPLSPDSRASSLSSESSPPKAMKXQAPARETCVE 392
Db 7 AAQATPSHDAKGGG--SSTVQ-----RSKSPS-----LRAQVKETCAA 42

QY 393 CQKTYPMERLLANQVPHISCFRCVYCNKLSLGTYSASLHGRYCKPHFNQLFKSKGNV 452
Db 43 CQKTYPMERLVADKLIFHNSCFCKCHTKLSLGSYAALHGEFYCKPHFQQLFKSKGNV 102

QY 453 DEGFGRPHKDLWASK 468
Db 103 DEGFGRKQHKELWAHK 118

RESULT 9
US-09-783-732-5
; Sequence 5, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; TITLE OF INVENTION: (EPLIN)
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09/783,732
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (0)...(0)
; OTHER INFORMATION: mutant sterol regulatory element binding protein 2
US-09-783-732-5

Query Match 7.2%; Score 281; DB 11; Length 52;
Best Local Similarity 96.2%; Pred. No. 5.7e-12;
Matches 50; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 390 CVCQKTYPMERLLANQVPHISCFRCVYCNKLSLGTYSASLHGRYCKPH 441
Db 1 CVGCQKTYPMERLLANQVPHISCFRCVYCNKLSLGTYSASLHGRYCKPH 52

RESULT 10
US-10-359-012-6
; Sequence 6, Application US/10359012
; Publication No. US20030232419A1
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; APPLICANT: KOLODKIN, Alex L.
; APPLICANT: TERMAN, Jon R.
; APPLICANT: MAO, Tianyi
; APPLICANT: PASTERKAMP, Ronald J.
; APPLICANT: YU, Hung-Hsiang
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEP
; FILE REFERENCE: JHUI840-3
; CURRENT APPLICATION NUMBER: US/10/359,012
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: US 60/388,325


```

/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 41..43
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 100..102
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 216..218
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 471..473
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 507..509
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 531..533
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 591..593
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 656..658
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 801..803
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 812..814
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 815..817
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 876..878
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 888..890
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 939..941
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1060..1062
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1128..1130
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1129..1131
/ OTHER INFORMATION: potential

```

```

/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1135..1137
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1181..1183
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1208..1210
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 1249..1251
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 47..50
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 126..129
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 157..160
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 158..161
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 159..162
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 216..219
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 274..277
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 276..279
/ OTHER INFORMATION: potential
/ FEATURE:
/ NAME/KEY: PHOSPHORYLATION
/ LOCATION: 295..298
/ OTHER INFORMATION: potential
/ FEATURE:

```

Query Match 5.7% Score 223; DB 12; Length 1312;
 Best Local Similarity 19.4%; Pred. No. 3.5e-06;

Matches 177; Conservative 150; Mismatches 329; Indels 258; Gaps 42;

```

Qy 40 KYQAAETNMKRSNTENLSQHFRKGLTLVLKWKWNPGLGAESHTDLSRNSSTEIRH 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 EFQWALPEKVVKQCKECENVKE-----IKVKEENETEIKEIKMEEERN----IIP 449

Qy 100 RADHPAEVTSHAASGAKADQEEQIHP-----RSRLRSPPEALVQGRYPHIK---DGEDL 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 450 REEKPIED-----EIERKENIKPSLGSKKNLLESIPTHSDQKEVNIKKPEDNENL 500

Qy 152 KDRSTESKXWENCL-----GSRHEVEKSEISENTDAS 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 DDKDDDTTRVDESINTKVAAEBEAKSGDETKEDEDEDEEAESEEEEEDEDDDDN 560

Qy 185 GKIEKYNVPLNLKWMFEXGEPTQTKILRAQSRAS-----GRKISENSYSL 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 561 NEEEFECYPPGMKVQVRYGRGNQIQWYEASIKDSVGEVGLVLYHYCGWNVRYDEWIK 620

```


NAME/KEY: PHOSPHORYLATION
LOCATION: 471..473
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 507..509
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 531..533
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 591..593
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 656..658
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 801..803
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 812..814
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 815..817
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 876..878
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 888..890
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 939..941
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1060..1062
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1128..1130
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1129..1131
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1135..1137
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1181..1183
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1208..1210
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1249..1251
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION

LOCATION: 47..50
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 126..129
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 157..160
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 158..161
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 159..162
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 216..219
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 274..277
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 276..279
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 295..298

Query Match

Best Local Similarity 5.7%; Score 223; DB 15; Length 1312;

Matches 177; Conservative 150; Mismatches 329; Indels 258; Gaps 42;

QY	40	KYKAAEETNMKKRSNTENLSQHFRKGTTLVLKKKNENPGLGAESHTDSLRSNSTEIRH	99
Db	403	EFQALPEKVVNKQCKECENYKE-----IKVKEENETEIKEMKEERN-----IIP	449
QY	100	RADHPPEVTSHAASGAKAQDEEQTHP-----RSRLRSPPEALVQGRYPHK---DGEDL	151
Db	450	REEKPIED-----EIEKENIKPSLGSKKKLLLESIPTHSQEKEVNIKKPEDNENL	500
QY	152	KDHSTESKMMENCL-----GESRHEVEKSEISENTDAS	184
Db	501	DDKDDTTRVDESINIKVEAEKAKSGDETCKEEDDEDEAESEEEEEDEDEDDDN	560
QY	185	GKIEKYVPLNRLXMMPEKGEPTQTKILRAQSRAS-----GRKISENSYSL	231
Db	561	NEEEFECYPPGMKVQVRYGRGKQKMYEASIKDSVGEVLYLVHYCGWNVRYDEWIK	620
QY	232	DDLEIGPGQQLS-----SSTFDSEKNE-----SRNLELPRLSETSIK-----DRMA	272
Db	621	ADKIVRPADKVVPIKIRKKIKNKLDKEKDEKYSPPNCKLRSLKPPFPOTNPSPMVS	680
QY	273	KYQAAVSKQSSS-----TNYTNELKASGGEIKIHKMEKENVPPGPEVCITHQGEK	324
Db	681	KLDLTDKNSDTAHIKSIEITSILNGIQAESSAEDSEQEDERGAQDMNNGKESKIDH	740
QY	325	ISANENSLAVRSTPAEDDSRDSQVKSE-----VQCPVHPKP-----LSPDS--	365
Db	741	LNNRNDLI-----SKEQNSSSLLEENKVHADLVISKVSKSPERLKRKDIENVLSEDTY	795
QY	366	-----RASSLSSESSPPKAMK-KFOAPARETC--VECOKTVYP--MERLLANQOVPHIS	413
Db	796	BEDEVTKRGKVDVKDDTKSKSPQIKGKGRYCNTEELKTGSGKKEEAKNKE-----	850
QY	414	CFRCYSY-----NNKLSLGTYSALHGRIVCKPHEN	443
Db	851	-----SLCMENSSNSSSDEDEETKAKMTPTKYNGLNLEEKRSKSLRTTGFYSG-----FS	899

QY 444 QL-----FKSGNYDEGFCH---RPHKDLWAS-----KNENEILERPAAQANARETPHS 490
DB 900 EVAEKRIKLLNDSERLONSRADKQDVWSSIQOWPKKTLKELFS-DSDETAASPPH- 957
QY 491 PVEDA-----PIAKVGVLAASMEAKASSQEKEDKPAETKKLRI-----AWP----- 533
DB 958 PAPEEXVAEESXQVAEESCSPELEKPPVNVDSKPTEETVEVNDRAEFPSSGSN 1017
QY 534 ----RPT--ELGSS-----GSALKEGKMSKPKWPEDEIS--KPEVPEDVLD----- 574
DB 1018 SVLNTPTTTPESSSVTVTSGSQOSSVTVSEPLAPNOQBEVRISKETOSTIEVDVSAE 1077
QY 575 LKLLRRSSILKERSRPTTVAASFOSTSVKSGPKTVSPPIRKGMWSQSESVGGRVAERK 634
DB 1078 LQDLQSEGN-----SSPAGFASVSSSSNQPEBHP--EKACTGQKRVKDAQGGSSSK 1131
QY 635 QVENAKA-----SKNGNVGKTTWQNKESKGETKRSK-----EGHSLE--MENENL 679
DB 1132 QKSHKATVYNNKKKGKGTNSDSEELSAGESITKQPVKSVSTGMKSHSTKSPARTQSP 1191
QY 680 VENGADSDSDNSFLKQOSPQEPKSLNWS---SFVDN-IFAEFTTQNKSQDVELWEGE 735
DB 1192 GKCGKNGKDPD--LKPSNRLPKVYKWSFQMSDLENMTSAERITILQEKLOEIRKHYLS 1249
QY 736 VKEL-SVERQIKR 748
DB 1250 LKSEVASIDRRKR 1263

RESULT 13
US-09-824-574-7
; Sequence 7, Application US/09824574
; Publication No. US20030077800A1
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Moilanen, Anu-Maarit
; APPLICANT: Palvimo, Jorma J.
; APPLICANT: Järne, Olli A.
; TITLE OF INVENTION: ARIP4 Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 2476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-824-574-7

Query Match 5.7%; Score 222; DB 11; Length 2476;
Best Local Similarity 20.6%; Pred. No. 9.5e-06;
Matches 180; Conservative 124; Mismatches 295; Indels 274; Gaps 41;

QY 12 TSLSLVATAKELSLVNNKSSAIV--EISKY-----QAAEETNWEKK- 53
DB 551 SSVKLVSSKDSRGNISKVTAKRVKELFVLTVPVSLNSPIKGVDCQEVQEKNGRKS 610
QY 54 ---RSN-----TENLSQHFRTGLTVLKKWENPGLGAESHSDSLRNSTSEIRHRAHPP 105
DB 611 GVARESEKCRPEEISDH--ENNVTIL-----LESDLRSPRVKTTPLRRQTESNP 660
QY 106 AEVTS-HAASGAKAQDEQIHF--RSRLRSPPEALVQGRYPH-IXDGED--LKQHSSTESK 159
DB 661 AMSNDEESNGTMKEQKMSGPIRKQKENSADCATDNPKPKHVPAKQPVIGQONSQSD 720
QY 160 KNENCLGESRHEVEKSEISENTDAGKIEKYNVPLNRLKMMFEKGEPTQTILRAQSRSA 219
DB 721 EMLAVLKEASQMGHSS--SSDTD-----INEP-----QNMHKGK-----TGKDD 757
QY 220 SGRKISENSYSYLDLLEIGPGQLSSSTFSEKXESRNLELPLSETSI--KORVAKYQAA 277

DB 758 NGKRRKNS-----TSGSDFDTKKGS-----TETSIISKKRQNY--- 793
QY 278 VSKQSSNTYNTELKASGEIKIHQEQ-----KENVPPGPEVCITHQE--GEKISAN--- 328
DB 794 ----SESSNYDEL-----EREIKTMSRIGAAKSVPEKKEEDSSEDEKQKVVNDNGH 844
QY 329 ENSLAVRSTPAEDDGRDSQVK-----SEVQOPVHPKPLSPDRASSLSSESP 376
DB 845 ERAKTTQEGSSADDTGDTGEGROGSCSAGSIEKVRSGVFREMLCKPGVSSDGAERPS 904
QY 377 KAMKQFAPARETCVECOKTVYPMERLLANQOVPHISCFRCSYCNKLSLGTASYALHGR 436
DB 905 VKEENVNSPEDK--RVSKTKEKTKHLSROS----- 933
QY 437 YCKPHNOLFSGKNGYDEGFHRPHKDLWASKNENEILERPAAQANARETPHSPGVEDA 496
DB 934 ----RKGKGGSSDCTDRFPKKEQSDSESEGEK-----KQSRQRPGTGK-KKA 975
QY 497 PIKAVGVLAASMEAKASSQEKEDKPAETKKLRTAWPPPTLGGSSGSALEEGIKMSK--- 553
DB 976 POLKGETLKRQEWDSDD-----CTERL-----PEEEIG-----PFSKGIKOSKTD 1019
QY 554 ----PKWPPDEISKPEVPEDVD-----LDLAKLRRSSSLKERSRFTVAA 595
DB 1020 AGGEKKGKWKDKCEKKEELSDSDVKLPFGKGDSCDSEDKTRNRVSLREKKR-FSLPA 1078
QY 596 SFQTSVSKPTVPPIRKGMSEQSESVGGRV-----AERKQ 635
DB 1079 ----KSPG-----KRPECSSDTEKSLKGCCDSTEKPKPRIDLRRERNSSKNT 1125
QY 636 VENAKSKNGNVGKTTWQNKESKGETKRSKEGHSLEMEENL-----VENGA 684
DB 1126 KEVSASSSSDAEGSSSDNKQKQRTSAKKKTGNTKEKKNLSRATPKRKQVDITSSSS 1185
QY 685 DSDEDDNSFLKQOSPQEPKSLNWSFVDNTPAEFTTQNKSQDVELW----- 732
DB 1186 DIGDDQNSAGESSDEQKI---KPVTEVLVPSHTGFCQSSGDEALSKSVPTATVDDDD 1242
QY 733 ---EGEVVVELSVEEQIKRNRYYDED--EDEE 759
DB 1243 DNDPENRIAKKMLLEE-IKANLSSDEGSSDDE 1274

RESULT 14
US-10-359-012-8
; Sequence 8, Application US/10359012
; Publication No. US20030232419A1
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; APPLICANT: KOLODKIN, Alex L.
; APPLICANT: TERMAN, Jon R.
; APPLICANT: MAO, Tianyi
; APPLICANT: PASTERKAMP, Ronald J.
; APPLICANT: YU, Hung-Hsiang
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEP
; TITLE OF INVENTION: AND METHODS OF USING THE SAME
; FILE REFERENCE: JHU1840-3
; CURRENT APPLICATION NUMBER: US/10/359,012
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: US 60/388,325
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/384,302
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/354,178
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 4723
; TYPE: PRT
; ORGANISM: Drosophila
US-10-359-012-8

Query Match 5.6%; Score 221; DB 12; Length 4723;
Best Local Similarity 17.4%; Pred. No. 2.7e-05;
Matches 184; Conservative 137; Mismatches 352; Indels 386; Gaps 37;

QY 15 SLRVAKELSLVNNKKSATVETPS-----KYOKAAEETNMKKNSNTLSQHFPRKGT 68
Db 646 SLDLTELE-SRIWLNLYDQICDLFRGEIPHIKHPKMDFDLQKRYRINHQAQDFSK-- 702
QY 69 LTVLKKWENPGLGAESHTSLNSSTEIRHAD-----HPPAEVTSHAASGAKA----- 118
Db 703 LLATPKAKGPMQDAVDIPITVORRSVLEERAKQRRHEQLLNIIGGAAGAAGVAGSG 762
QY 119 ----DQEQIHRP-----SRLSPPEALVQGRPHIKDGEDLKDHSSTESKXNENCLGESRHE 171
Db 763 TGITTCQNDTPRSKKRQVDTANIEERQQLQIEENRQERMKRQORC-----HQ 817
QY 172 VEKSEISENTASGKIEKYNVPLNRLLKMFKEGPTQTKILRAQSASGKISSENSYSL 231
Db 818 TQNFYKSLQLLAG-----KLLREGGE-----AGVAEDGTTP 849
QY 232 DDLEIGPGQLSSSTFSEKXESRNL-----ELPRUSETSIKDRMAKO----- 275
Db 850 EDYSIFLYROQAFVENDRVKDLERKLLFPDRERGDIPSALPRTADEQFSDRIKMEQRM 909
QY 276 -----AAVSKQSSSTNYTNELKAS-----GGEIKIKHKEQXENVPPEVCV 317
Db 910 GRGLGDKKPKDLMRAIGKIDNDNWNVREIEKIELSKKTEIHGPKREKVPKWSK---- 966
QY 318 THOGEKISANENS LAVRSTPAEDDSDSQVSEVQPVH--PKPL----- 361
Db 967 ----EQFQARQHKM--SKPORQDSREAEKFKDIDQITIRNLDKQLKEGHNLDVGERGN 1018
QY 362 -----SPQSRAS-----LSESSPPKAMKKFOAPARETCVEQ 394
Db 1019 KVASIAQOGKKDEANSDEKNAGSNATTNTNTVTPKSSSKVALAFKQAAASEKCRFC 1078
QY 395 KTVYMERLLANQOVFHISCFRCSYCNKLSLGTVA-----SLHGRYCKPHF----- 442
Db 1079 QTYPNKETTVEGLVLRNCLKCHCHTNLRUGGYAFDRDDQGRLYCTQHFLPPKPLP 1138
QY 443 ---NOLFKS-----KGNVDEGFGRPHKDL----- 464
Db 1139 QRTNARKASAAQAPASPAVPTAGSVPTAAATSEHMDTTPRDQVDVLLTOTSANASADM 1198
QY 465 -----WA-----SKVNEEILERAQLANARETHPSGVEDAPIAKGVGLA 505
Db 1199 SDDEANVIDEHWSGRNFLPESNDSQSELSSDESDETSDEMEFEADDSFFGAQTLQL 1258
QY 506 AS-----MEAKASSQOEKEDKPAETKKLRIAWPP-----PT-- 536
Db 1259 ASDWIKQYCEDDSDDDFYDSEDDGDDTEGEFEKARELRQEVRLQPLPANLPTDT 1318
QY 537 -----ELGSSGSALEEGIKWSKPKWPPPEDEISKP----- 565
Db 1319 ETEKULNVNDKNENMADGSSLSKSGNSFESARQSPSTPLTPTRVEMEQLERDAPRKFSE 1378
QY 566 -----EVPEDVDLKK-----LRRSSSLKER-----SR 589
Db 1379 IEAISEKLYHMNMVQNKDLEVLAKENLVKSGILRLKTLKEKLAENAAIAAGOKVTPT 1438
QY 590 PFTVAASFQSTS---VKSPKTVSP-----IRKGMWSMEQSEESV 626
Db 1439 PSATAPGLQPKSFKDEKFEKVPVSPQVVEPKPKPVDFNLDLXPKRP-NFEERPKQL 1497
QY 627 GQVRAERKOVENAKSKNGNKGKTTWQNKESKGETG---KRSKEGHSLEWENLV-- 680
Db 1498 -PRPSLKPPQPKPGKSGSTNVRS--NSLKGNASNGSPVKKAPVSNNSKMQIEGILDT 1554
QY 681 -----ENGADSDDD-----NSFLKQ-----QSP 699
Db 1555 LRKIQSNSSDDQDMVDVEDVERKENKELNSKLKIEIQASSFAGTMDHIKSLQTMPTVSA 1614
QY 700 QEPKSLNWSFVDNTFAEFTTQNKSQDVELWEGEVVK 738

Db 1615 QAPPSMDLSKFPNOKQEKSTSTSNKNQVTLKQVNLAK 1653

RESULT 15
US-10-359-012-4
; Sequence 4, Application US/10359012
; Publication NO. US20030232419A1
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; APPLICANT: KOLODKIN, Alex L.
; APPLICANT: TEERMAN, Jon R.
; APPLICANT: MAO, Tianyi
; APPLICANT: PASTERKAMP, Ronald J.
; APPLICANT: YU, Hung-Hsiang
; TITLE OF INVENTION: MOLECULES INTERACTING WITH CASL (MICAL) POLYNUCLEOTIDES, POLYPEP
; FILE REFERENCE: JHU1840-3
; CURRENT APPLICATION NUMBER: US/10/359,012
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: US 60/386,325
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/384,302
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/354,178
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1633
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-359-012-4

Query Match 5.4%; Score 213; DB 12; Length 1633;
Best Local Similarity 21.0%; Pred. No. 2.2e-05;
Matches 166; Conservative 92; Mismatches 284; Indels 248; Gaps 34;

QY 46 EETNMKKRSNTENLSQHFRKGL--TVLKKWENPGLGAESHTDSDRNSTETIRHRAH 103
Db 708 EGGNQNVKSWANQLLAKFEESTENPSLMKQERKVSIG-----KPVLCSSSGPPVHSCCP 763
QY 104 PPAEVTSHAASGAKADQEEQIHPRSLRSPPEALVQGRPHIKDGEDLKDHSSTESKXNEN 163
Db 764 KPEEATSPSPPLK-----RQPSVVVTG--HVL--RELKQVSAGSE----- 801
QY 164 CLGSRHEVEKSEISENTASGKIEKY-NVPLNRLLKMFKEG-----EPTQTKILRAQSR 217
Db 802 CLSRPWRARAKSDIQ-----LGGTENFATLPSTPRQAALSGVLWRLLQOVVEEKILQRAQ 856
QY 218 SASGRK-----ISENSYSLDDLEIGPCQLSSSTFSEKXESRNLRLPRUSETSIKDRMA 272
Db 857 NLANREFTTKNIREKAAHLASM-FGHG-----DPPQNK----- 888
QY 273 KYQAAVSKQSSSTNYTNELKASGGEIKIKHKEQXENVPPEVCITHQGEKISANENSL 332
Db 889 ----LLSKGLSHTH-----PPSPSRPLSPOP----- 911
QY 333 AVKSTPAEDDSDRSQVSEVQQP-----VHPKPLSP----- 363
Db 912 AASSPSTVDSA-SPARKEKKSFGHFFHSHLRTVHPQLTVGKVSAGGIGAAAABVLNLY 970
QY 364 --DSRASSLSESPPKAMKK---FOAPARETCVEQKTVYMERLLANQOVFHISCFRCS 418
Db 971 MNDHRPKAQATSPDLSEMRKSFPLNLGSGDTCYFKKRVYVMERLSAEGHFFHRECFRCS 1030
QY 419 YCNKLSLGTVA--SLHGRYCKPHFNQLFKXGNVDEGFGHPRPHKDLWASKNENBEILE 476
Db 1031 ICATTIRLAAYTDCDEGKFCYCKPHF---IHCKTN-----SKQRRRAELKQOREEAT 1081
QY 477 RPAQLANARETHPSGVEDAPIAKGVGLAA-----SME 509
Db 1082 WQEQEAPRDTPTTE---SSCAVAAGTLEGSPPVHFSLPVLHPLLGLMLDWNDSIPESVH 1138

```
QY 510 AKAS---SQEKED-----KPAETKKL-RIAWPP-PTLGGSGSALKEGKMSKPKWP 557
Db 1139 LKAGERISQKSAENGGRVLPVPRLLLPRAAGEPLPTORCAQE-----KMGTPAEQ 1191
QY 558 PEDEISKPEVPEDVDLDLKKLRRS-----SSLKERSRPFTVAASFQSTSVKS 604
Db 1192 AQGERNVPPKGPLRLIANAIRRSLEPLLSNSEGKKAWAKQESKTLPTQACTRSFS--- 1248
QY 605 PKTVSPPIRKGSMSQSESVGGRVAERKQVENAKSKKNGNVGKTTWQNKESKGETGK 664
Db 1249 -----LRKTNKDGDOHSPG-----RQSSAFPPDPALRTHSLPNRPSKVFPAL 1294
QY 665 RSKEGHSLE-----MENENLVENGADSDDDNSFLKQSQSPQEPKSLNWSSFYDNTFAEEF 719
Db 1295 RSPPCKSIEEVTLLKYSLOENFPDA-----SKPPKXRSISLPSLSRLKDKSPFESF 1345
QY 720 TTQNKSQDV 729
Db 1346 LOESRQRKDI 1355
```

Search completed: January 6, 2004, 09:50:44
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 6, 2004, 20:07:07 : Search time 7226 Seconds
(without alignments)
4297.039 Million cell updates/sec

Title: US-09-890-549-4
Perfect score: 3927
Sequence: 1 MESSPFNRQWTSLSRVTA.....LSVBEQIKENRYDEDEEE 759

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2885711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2.1/USFTO.spool/US09890549/runat_06012004_094752_19821/app_query.fasta_1.903
-DB=GenEmbl -OPMT=fastap -SUFFIX=p2n.rge -MINWATCH=0.1 -LOPCU=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09890549 @CNC 1.1 4617 @runat_06012004_094752_19821 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:
1: gb.ba.*
2: gb.hts.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3927	100.0	3655	9	AF198454 Homo sapi
2	3927	100.0	3664	6	AX086386 Sequence
3	3927	100.0	3664	9	AL136911 Homo sapi
4	3899	99.3	2536	9	AK000372 Homo sapi
5	3640.5	92.7	3550	9	AF198455 Homo sapi
6	3417.5	87.0	2164	9	BC001247 Homo sapi
7	3416	87.0	3277	9	AK000335 Homo sapi
8	2970	75.6	3232	10	BC031490 Mus muscu
9	2967	75.6	2370	10	AF307845 Mus muscu
10	2710	69.0	3997	10	AF307844 Mus muscu
11	2679	68.2	1754	9	AF218025 Homo sapi
12	2465	62.8	2207	6	BD160117 Primer fo
13	2465	62.8	2207	9	AX023649 Homo sapi
14	2460	62.6	2793	6	AK096172 Homo sapi
15	2458	62.6	2783	6	AX017487 Sequence
16	2458	62.4	3348	6	BD135177 Human nuc
17	2449	62.4	3348	9	AF157325 Homo sapi
18	2376.5	60.5	2667	9	AK000057 Homo sapi
19	1749.5	44.6	115345	9	AC008147 Homo sapi
20	1519	38.7	4425	10	CGU22818 Cricetus
21	1450	36.9	1021	6	AX012180 Sequence
22	1421	36.2	2121	9	BC010664 Homo sapi
23	1324	33.7	195301	2	AC134548 Mus muscu
24	1209	30.8	237661	2	AC114454 Rattus no
25	1128	28.7	2631	5	AF307846 Dario rer
26	1127	28.7	2678	5	BC047797 Dario rer
27	1103	28.1	777	11	G56637 SHGC-102035
28	881	22.4	641	6	AX011639 Sequence
29	657.5	16.7	2264	9	AK097519 Homo sapi
30	641	16.3	3287	9	AB097518 Macaca fa
31	617.5	15.7	5264	9	AL832452 Homo sapi
32	615.5	15.7	4548	9	HSM803760 Homo sapi
33	614.5	15.6	4796	9	HSM803292 Homo sapi
34	614.5	15.6	5175	9	HSM803668 Homo sapi
35	614.5	15.6	5216	9	HSM803318 Homo sapi
36	613.5	15.6	4885	9	HSM803277 Homo sapi
37	612.5	15.6	4815	9	HSM803690 Homo sapi
38	609.5	15.5	4225	9	HSM803683 Homo sapi
39	608.5	15.5	2710	6	AX714335 Sequence
40	608.5	15.5	2710	9	AK056582 Homo sapi
41	606	15.4	413	6	AR270117 Sequence
42	602.5	15.3	6524	9	HSM804604 Homo sapi
43	596.5	15.2	3825	6	AX375615 Sequence
44	576	14.7	3516	4	AY246700 Sus scro
45	575.5	14.7	195516	2	AC135476 Rattus no

ALIGNMENTS

AF198454
LOCUS AF198454 3655 bp mRNA linear PRI 10-JAN-2000
DEFINITION Homo sapiens epithelial protein lost in neoplasm beta (EPLIN) mRNA,
complete cds.
ACCESSION AF198454
VERSION AF198454.1 GI:6685006
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3655)
AUTHORS Maul, R.S. and Chang, D.D.
TITLE EPLIN, epithelial protein lost in neoplasm
JOURNAL Oncogene 18 (54), 7838-7841 (1999)
MEDLINE 20087188
PUBMED 10618726
REFERENCE 2 (bases 1 to 3655)
AUTHORS Maul, R.S. and Chang, D.D.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los
Angeles, CA 90095, USA
FEATURES
source
1. 3655
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q13"
1. 3655
/gene="EPLIN"
102..2381
/gene="EPLIN"
/codon_start=1
/product="epithelial protein lost in neoplasm beta"
/protein_id="AAF23755.1"
/db_xref="GI:6685007"
/translation="MSSPNRRQWTSLSRVTAKELSLVNKNKSAIVEIFSKYOKA
AETNMEKKSNTENISQHPKGTGLTVLKKWENPGGAESHTSLRNSSTIRHRAD
HPEAVTSHAAGKADQEEQIHPRLRLSPPEALVQGRYPHIKQGEDILKQSTESKK
MENGLGESRHEVSEKSEITDASGKIEKYNVPLNLKWMFKGEPOTQKILRAOSRS
ASGRKISSENSYLDLEIGPGQLSSSTFSEKNESRNLELPLRSITSIKRWAKYQA
AVSKQSSNTYNELKASGGEIKIHWKQKENVPPGECVITHQSEKISANENSLAV
RSTPAEDSDSQKSEVQVPFVPPKLSPSRASSLSSESPPKAMKKFOAPARETCVE
CQKTYPMERLAKQVHFHISPCRSYCNKKLSLGYASLHGRYICKPHFNQLFKSG
NYDEGFGHDKDLWASKNENELIRPAQLANARETHPSGVEDAPIAKVGLAASM
EAKASQOEEDHPATKKLRIAWPPTELGGSSGSALEGIKMSKPKWPEDEISKPE
VPSVDLDLKLRLSSSLKERSRFTVAASFQSTSVKSPKTVSPPIRKWNSQSEGE
SVGRVAERQVENAKASKNGVNTWQNKESGETGRKSKESHLEMNENLVEN
GADSDDDNLSFLKQSPQEPKSLNWSFSVDNTFAEEFTTQNKSDQVELWEGEVVKEL
SVBEQIKRNRYYDEDEDEE"
3632..3637
polyA_signal
/gene="EPLIN"
BASE COUNT 1197 a 737 c 810 g 911 t
ORIGIN
Alignment Scores:
Pred. No.: 3,34e-206 Length: 3655
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-890-549-4 (1-759) x AF198454 (1-3655)
QY 1 MetGluSerSerProPheAsnArgGlnTTPThrSerLeuSerLeuArgValThrAla 20
Db 102 ATCGAATCATCTCCATTATATAGACGGCAATGACCTCACTATCATTTAGGCTAACAGCC 161
QY 21 LysGluLeuSerLeuValLeuLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 162 AAGTTCAGGACCTCGCAAGAGAGACCTCGGTGGAATGTGAGAGACAGCTATTCCTCAAG 221

QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 222 TACCAGAAAGCAGCTGAAGAAACAACATGGAGAAGAGAGAGTAACACCCAAAATCTC 281
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 282 TCCAGCAGCTTTAGAAAGGGGACCTGACTGTGTTTAAAGAGAGTGGGAGAACCCAGG 341
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 342 CTGGGAGCAGAGCTCTCACACAGACTCTCTTACGGAACAGCAGCAGCTGAGATTAGGCACAGA 401
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaIleSerGlyValAlaLysAlaAspGln 120
Db 402 CGAGACCATCTCTCTGCTGAAGTACAGGCACGCTGCTTCTGGAGACCAAGCTCAGCCAA 461
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 462 GAAGAACAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAAGCCTCTGTTCAAGGT 521
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 522 CGATATCCCCACATCAAGACCGTGAGATCTTAAAGACCACCTCAACAGAAAGTAAAAA 581
QY 161 MetGluAsnCysLeuGlyLysSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 582 ATGGAAATTTGCTAGGAGAAATCCAGGCATGAAGTAGAATAATCAGAAATCAGTGAAGAAC 641
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db 642 ACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGAGCTTAAGATGATG 701
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
Db 702 TTTGAGAAGGTGAACCAACTCAACTAAGATTCTCCGGGCCCAAGCCGAGTCAAGT 761
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
Db 762 GGAAGAGAGATCTCTGAAAACAGCTATTCTTAGATGACCTGGAAATAGGCCAGGTGAG 821
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db 822 TTGTCATCTTCTACATTTGACTCGGAGAAAAATGAGAGTAGACGAAATCTCGAAGCTTCA 881
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db 882 CGCCTTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAGGCAGCTGTGTCCAAA 941
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
Db 942 CAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGGCGAAATCAAAAT 1001
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db 1002 CATATAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCTCTGAGGTCTGCTCACCCTCAG 1061
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 1062 GAAGGGGAAAAAGATTCTGCAAAATGAGAAATAGCTGCGAGTCCGTTCCACCCCTGCCGAA 1121
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
Db 1122 GATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCCTGTCCATCCCAAGCCA 1181
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1182 CTAAGTCAGATTCCAGAGCCCTCCAGCTTTCTTGAAAGTTCTCTCTCCCAAGCAATGAAG 1241
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValThrProMet 400
Db 1242 AAGTTTCAGGCACCTCGCAAGAGAGACCTCGGTGGAATGTGAGAGACAGCTATTCCTCAATG 1301

401 GluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
1302 GAGCGTCTCTGGCCACACAGCAGTGTTCACATCAGCTGCTTCCGTTCTCTATTGC 1361
421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
1362 AACAAACAACCTCAGTCTAGGAACATATGCACTTTACATGGAAGATCTATTGTAAGCCT 1421
441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
1422 CACTTCAATCAACTCTTTAAATCTTAAGGCAACTATGATGAAGCTTTGGCCACAGACCA 1481
461 HisLysAspLeuTyrAlaSerLysAsnGluAsnGluGluLeuLeuGluArgProAlaGln 480
1482 CACAAGATCTATGGGCAAGCAAAATTAAGAAAGAGATTTTGGAGACCCAGCCAG 1541
481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
1542 CTTCCAAATGCAAGGGAGACCCCTCAGAGCCAGGGGTAGAGATGCCCTATTGCTAAG 1601
501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
1602 GTGGGTCTCTGGTGTCAAGTATGGAAGCCAGGCTCTCTCAGCAGGAGAGGAGAC 1661
521 LysProAlaGluThrLysLysLeuArgIleAlaTyrProProThrGluLeuGlySer 540
1662 AGCCAGCTGAACCAAGAGCTGAGATCGCTGGCCACCCCTGAACTGAACTGGAAGT 1721
541 SerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTyrProGluAsp 560
1722 TCAGGAAGTCCCTGGAGGAGGATCAAAATGTCAAGGCCAAATGGCTCTCTGAAGAC 1781
561 GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
1782 GAAATCAGCAAGCCGAAATCTCTGAGATGTGATCTAGATCTGAAGAGCTAAGACGA 1841
581 SerSerSerLysLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
1842 TCTTCTTCACTGAAGGAAGAAAGCCGCCATTCACGTAGACCTTCATTTCAAGACCC 1901
601 SerValLysSerProLysThrValSerProPheArgLysGlyTyrPheSerMetSerGlu 620
1902 TCTGTCAAGAGCCCAAAATCTGTCCCACTATCAGGAAGGCTGGAGCATGTGAGAG 1961
621 GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
1962 CAGAGTGAACAATCTGTGGGTGAAGAGTTGCAGAAAGGAACAAGTGGAAATGCCAAG 2021
641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
2022 GCTTCTAAGAAAGATGGGAATGTGGGAAACCAACCTGGCAAAACAAAGATCTAAAGGA 2081
661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
2082 GAGACAGGGAAGAGAGTAAGGAAGGTCAAGTTTGGAGATGGAATGAGATCTTGTA 2141
681 GluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLysGlnGlnSerProGln 700
2142 GAAATGGTGCACACTCCGATGAAGATGATACAGCTTCTTAAACAAACAATCTCCACAA 2201
701 GluProLysSerLeuAsnTrpSerPheValAspAsnThrPheAlaGluGluPheThr 720
2202 GAACCAAGTCTCTGAATGGTGCAGTTTGTAGACAACACCTTCTGCTGAAGAAATCACT 2261
721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGluValValLysGluLeu 740
2262 ACTCAGAAATCAGAAATCCCAAGGATGTGGAACCTGGGAGGGAAGTGTGCAAGAGCTC 2321
741 SerValGluGluGlnLysLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
2322 TCTGTGAAGAACAGATATAAGAGAAATCGGTATTATGATGAGATGAGATGAAGAG 2378

RESULT 2

AX086386
LOCUS AX086386 3664 bp DNA linear PAT 09-MAR-2001
DEFINITION Sequence 338 from Patent WO0112659.
ACCESSION AX086386
VERSION AX086386.1 GI:13275951
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Wienann, S.
Human dna sequences
Patent: WO 0112659-A 338 22-FEB-2001;
German Human Genome Project (DB)
Location/Qualifiers
1. 3664
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 1210 a 734 c 809 g 911 t
ORIGIN
Alignment Scores:
Pred. No.: 3.35e-206 Length: 3664
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-890-549-4 (1-759) x AX086386 (1-3664)
QY 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 94 ATGGAATCATCTCCATTTAATAGACGGCAATGGACCTCACTATCATTCAGGGTAACAGCC 153
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 154 AARGAACTTTCTCTGTCTCAAGAACAAAGTCATCGGCTATTGTGGAATATTCTTCCAG 213
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 214 TACCAGAAAGCAGCTGAAGAAACAACATCGAGAGAGAGAGTAACACCGAAATCTC 273
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 274 TCCAGCAGCTTTAGAAAGGGGACCTGACTGTGTTAAAGAAAGAGTGGGAGAACCCAGGG 333
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 334 CTGGAGCAGAGTCTTCACACAGACTCTCTACGGAACACAGCAGCTAGATTAGGCACAGA 393
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
Db 394 CGAGNCCATCTCTCTGTAAGTAGAACAGCCACCTGCTCTGAGGCCAAAGCTGACCAA 453
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 454 GAAGAACAAATCCACCCAGATCTAGACTCAGTCACTCTCTGAGGCCCTCTGTCAGGGT 513
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 514 CGATATCCCAACATCAAGAGCGGTGAGGATCTTAAAGACCACTCAACAGAAAGTAAAAAA 573
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 574 ATGGAATTTCTTAGAGAAATCCAGCATCAAGTAGAAAAATCAGAAATCAGTGAAGAAC 633
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db 634 ACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAGATGATG 693

QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuAArgAlaGlnSerArgSerAlaSer 220
DB 694 TTTGGAAGAGTGAACCACTCAAACTAGATTCTCCGGGCCCAAGCCGAAGTGCAGT 753
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
DB 754 GGAAGGAAGATCTCTGAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCAGGTCAG 813
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
DB 814 TTGTCTATCTTCTACATTTGACTCGGAGAAATAGAGATAGACGAATCTGGAACTTCCA 873
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
DB 874 CGCTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAGGAGCTGTGTCCAAA 933
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
DB 934 CAAGCAGCTCAACCACTATACAAATGAGCTGAAGCCAGTGGTGGCGAAATCAAAATT 993
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
DB 994 CATAAATGGAGCAAAAGGAGATGTGCCCCAGTCTCGAGGTCTGCATCACCCTATCAG 1053
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
DB 1054 GAAGGGAAAGATTTCTGCAATGAGAAATAGCTGGCAGTCCGTTCACCCCTGCCGAA 1113
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
DB 1114 GATGACTCCCGTACTCCAGCTTCAAGAGTCAAGTCAAGCCTGCTCCATCCCAAGCCA 1173
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
DB 1174 CTAAGTCCAGATTCAGAGCTCCAGTCTTTCTGAAAGTCTCCTCCCAAGCAATGAAG 1233
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
DB 1234 AAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGAATGTCAAGAGACAGTCTATCCAATG 1293
QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
DB 1294 GAGCGTCTCTTGGCCCAACCAAGCTGTTCACATCAGCTGCTTCCGTGCTCTATTGC 1353
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
DB 1354 AACCAACAACTCAGTCTAGGAACATATGCTATCTTACATGGAAGATCTATTGTAGCCT 1413
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
DB 1414 CACTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGCTTTTGGGACAGACCA 1473
QY 461 HisLysAspLeuTrpAlaSerLysAsnGluGluIleLeuGluArgProAlaGln 480
DB 1474 CACAAGGATCTATGGGACGCAAAATGAAACCAAGAGATTTTGGAGACCAAGCCAG 1533
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
DB 1534 CTTGCAAAATGCAAGGAGAGACCCCTCAGAGCCAGGGTAGAGATGCCCTATTGCTAAG 1593
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysGluAsp 520
DB 1594 GTGGGTGTCTGGTGAAGTATGGAAGCCCAAGGCTCCTCTCAGCAGGAGAGGAAGAC 1653
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
DB 1654 AAGCCAGCTGAAACCAAGAGCTGAGATCGCTGGCCACCCCTCAGTGAACCTGGAAGT 1713
QY 541 SerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTrpProProGluAsp 560
DB 1714 TCAGGAAGTGCCTTGGAGGAGGATCAAAATGTCAAAGCCCAAAATGGCCCTCTCTGAGAC 1773
QY 561 GluLysSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580

DB 1774 GAATCAGCAGCCCGAAGTCTCTGAGGATGTCGATCTAGATCTGAAGAAGCTAAGACGA 1833
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
DB 1834 TCTTCTTCACTGAAGAAAGAGCCGCCATTCACCTGTAGCAGCTTCATTTCAAAGCACC 1893
QY 601 SerValLysSerProLysThrValSerProProIleArgLysGlyTyrSerMetSerGlu 620
DB 1894 TCTGTCAAGAGCCCAAAACTGTGTCCCACTATCAGGAAAGGCTGGAGCATGTCAGAG 1953
QY 621 GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
DB 1954 CAGAGTGAAGAGTCTGTGGTGAAGAGTTCAGAAAGAAACAAGTGAATGCAAG 2013
QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
DB 2014 GCTTCTAAGAGATGGGAATGTGGGAAAAACAACCTGGCAAAACAAAGATCTAAAGGA 2073
QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
DB 2074 GAGACAGGGAAGAGTAAAGAGGTGATAGTTTGGAGATGGAGATCAGAATCTTGTA 2133
QY 681 GluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLysGlnSerProGln 700
DB 2134 GAAATGGTGCAGACTCCAGATGAAGATGAACAGCTTCTCAAAACAACATCTCCACA 2193
QY 701 GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPheThr 720
DB 2194 GAACCAAGTCTCTGATTTGGTTCGAGTTTTTGTAGACAACACCTTTGCTGAAGATTCAT 2253
QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyValValLysGluLeu 740
DB 2254 ACTCAGAAATCAGAAATCCAGGATGTGGAACCTCTGGAGGAGAGAGTGTCAAGAGCTC 2313
QY 741 SerValGluGluGlnLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
DB 2314 TCTGTGGAAGAACAGATAAAGAGAAATCGTATTATGATGAGATGAGGATCAAGAG 2370

RESULT 3
HSM800164 3664 bp mRNA linear PRI 03-AUG-2001
LOCUS Homo sapiens mRNA; cDNA DKFZp58611918 (from clone DKFZp58611918);
DEFINITION complete cds.
ACCESSION AL136911
VERSION AL136911.1 GI:12053316
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 3664)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2001) MIPS, Am Klopferspitze 18a, D-82152
Martinsried, GERMANY
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp58611918) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/
LOCATION/Qualifiers
FEATURES
1. 3664
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="12"
/clone="DKFZp58611918"
/tissue_type="uterus"

```

/clone_lib="586 (synonym: hutel). Vector pSport1; host
DH10B; sites NotI + SalI/MluI"
/dev_stage="adult"
1. 3664
/gene="DKFZp58611918"
94. 2373
/notes="DKFZp58611918"
/notes="similarity to transcription factor SF3"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB66845.1"
/db_xref="GI:12053317"
/db_xref="SPRMBL:Q9UHB6"
/translation="MESSPNRRQWTSLSRVTAKELSLVNKNKSSAIVEIFSKYOKA
AETNMEKRSNTENSQHFQKGTLTVLKKKWNPGLSAHSHTDSLRNSSTETIRHRAD
HPAEVTSAAAGAKAQDEQIHPRSRLRSPPEALVQGRVPHIKDGEDLKHSTESKK
MENCLESREVEKSEITSENDSAGKIEKYNVPLNLRKMFKEGETQIKILRAQSR
ASGRKSENYSJDDLEIGQJSSSTDSKNESENUELPELSLETQIKILRAQSR
AVSQSSSTNYNELKASGGEIKHKWEQENVPQPEVCITHOEGEKISANENSLAV
RSTPAEDSDSVQKEVQVPHKPLSPDSRASISSESPPKAMKKFOAPARETCVE
CQTVVPMERLLANQVHFISCFRCYCNKLSLGTVASLHGRIYCKPHFNOLFKSKG
NYDEGFHRPHKDLWASKNEBILERPOLANARETTPHSPGVEDAPIAKVGLAASM
EAKASQOEEDKPAETKILRIAMPPTTELSSGSALBEGIKMSKPKWPPDEISKPE
VPEDVDJLKKLRSSLSKERSRPFVTAASFQTSVSKSPKTSVPPIRKGMSEQSEE
SVGVRVAERQVENAKASKKNGVGTWQNKESKETGRKSEGHSLMEWENLVEN
GADSDDEDSFLKQOQSPQEPKSLNWSFVNTFAEBFTQNKQSDVELWEQEVKEL
SVFEQIKRNRYYDEDEDE"
polyA_signal 3637 3642
/gene="DKFZp58611918"
3647
polyA_site 3647
/gene="DKFZp58611918"
1210 a 734 c 809 g 911 t
BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 3,35e-206 Length: 3664
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-890-549-4 (1-759) x HSM800164 (1-3664)

QY 1 MetGluSerSerProPheAsnArgArgGlnThrTrpThrSerLeuSerLeuArgValThrAla 20
Db 94 ATGGAATCATCTCCATTTAATAGACGGCAATGACCTCACTATCATTTGAGGGTAACAGCC 153
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 154 AAAGAACTTCTCTTGTCAACAGAACAAAGTCAATCGGCTATTGTGGAAATATTTCTCCAA 213
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 214 TACCAGAAAGCAGCTCAAGAAACAAACATGGAGAGAGAGAGTAACACCCGAAATCTC 273
QY 61 SerGluHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 274 TCCACACCTTTAGAAAGGGGACCTGACTGTGTTAAAGAGAGTGGGAGAACCCAGGG 333
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 334 CTGGGAGCAGAGTCTCACACAGACTCTTACCAGGACAGCAGACTGAGATTAGGCACAGA 393
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyValAlaLysAlaAspGln 120
Db 394 GCAGACCATCTCTCTGCTGAAGTACAGACCAAGCTGCTTCTGGAGCAACAGCTGACCAA 453
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 454 GAAGAACAAATCCACCCAGACTAGACTCAGCTCACCTCTCTGAGAGCCCTCGTTGAGGGT 513
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160

```

```

Db 514 CGATATCCCCACATCAAGGACGGTGAGGATCTTAAGACCATCTAACAGAAAGTAAAAA 573
QY 161 MetGluAsnCysLeuGlyCysSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 574 ATGGAAAAATTTCTAGGAGATCCAGGATGAAGTAGAAAAATCAGAAATCAGTGAAC 633
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db 634 ACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTCAACAGGCTTAAGATGATG 693
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
Db 694 TTTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGGAAGTCAAGT 753
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
Db 754 GGAAGGAAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCACAGGTGAG 813
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db 814 TTGTCTATCTTCTACATTTGACTCGGAGAAAAATAGAGTAGACGAAATCTGGAACCTTCCA 873
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db 874 CCCTCTCAGAAACCTCTATAAAGATCGAATGGCCAAAGTACCGAGCAGCTGTGTCCAAA 933
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
Db 934 CAAAGCAGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAAT 993
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db 994 CATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCTCTGAGGCTGTGATCACCCTCAG 1053
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 1054 GAAGGGGAAAAAGATTTCTGCAATAGCAATAGCTGCGAGTCCGTTCCACCCCTGCCGAA 1113
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
Db 1114 GATGACTCCCTGACTCCAGGTAAAGAGTGAGGTTCACAGCCTGTCCATCCCAAGCCA 1173
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1174 CTAAGTCCAGATTCAGAGCCTCCAGCTTTTCTGAAAGTTCTCTCCCAAGCAATGAG 1233
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
Db 1234 AAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTCAGAGACAGCTATATCCAATG 1293
QY 401 GluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
Db 1294 GAGCGTCTCTTTGGCCCAACCCAGCAGGTGTTTCCATCAGCTGCTTCCGTTGCTCTATTGC 1353
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
Db 1354 AACACAAACTCAGCTTAGGAACATATGCAATCTTTATCATGGAAGAAATCTATTGTAAGCCT 1413
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1414 CACTTCAATCACTCTTTAAATCTAAGGGCAACTATGATGAGGCTTTGGCAGACAGACCA 1473
QY 461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluIleLeuGluArgProAlaGln 480
Db 1474 CACAAGGATCTATGGGCAAGCAAAATGAAACCGAGAGATTTTGGAGAGACCCAGCCAG 1533
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1534 CTTGCAAAATCAAGGGGAGACCCCTCAAGCCAGGGGTAGAGATGCCCTATTGTTAAG 1593
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysGluAsp 520

```

Db	1594	GTGGGTGCTCGCTGGCTGCAAGTATGGAAGCCAGGCTCTCTCAGCAGGAGGAAGAC	1653
QY	521	LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer	540
Db	1654	AAGCCAGCTGAACCAAGAGCTGAGGATCGCTGGCCACCCCTGAACTTGGAGT	1713
QY	541	SerGlySerAlaLeuGluGluGlyLysMetSerLysProLysTrpProGluAsp	560
Db	1714	TGAGGAAGTGCCTTGGAGGAAGGATCAAAATGTCAAAGCCAAATGGCCCTCCTGAAGAC	1773
QY	561	GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLeuArgArg	580
Db	1774	GAAATCAGCAAGCCGGAAGTTCCTGAGGATGCGATCTAGATCTGAAGAGCTAAGACGA	1833
QY	581	SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr	600
Db	1834	TCTTCTTCACTGAGGAAGAAGCCGCCCAATCTACTGTAGCAGCTTCATTTCAAAGCACC	1893
QY	601	SerValLysSerProLysThrValSerProProIleArgLysGlyTrpSerMetSerGlu	620
Db	1894	TCTGTCAAGAGCCCAAAACATGTGTCCCACTATCAGGAAGGCTGGAGATCTCAGAG	1953
QY	621	GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys	640
Db	1954	CAGAGTGAGAGTCTGTGGTGGAGAGTTCGAGAAAGCAACAGTGGAAAATGCCAAG	2013
QY	641	AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly	660
Db	2014	GCTTCTAAGAGAAATGGGAATGTGGGAAACCAACCTGGCAAAACAAAGAAATCTAAAGGA	2073
QY	661	GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluLeuVal	680
Db	2074	GAGCAGGGAAGAGAGTAAAGAGGTCATATTGGAGATGGAGAAATGGAATCTTGTA	2133
QY	681	GluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLysGlnGlnSerProGln	700
Db	2134	GAAATGTGTGAGACTCCATGAGATGATTAACAGCTTCTTCAAACCAACATCTCCCAA	2193
QY	701	GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPheThr	720
Db	2194	GAACCAAGTCTCTGAATGTGCGAGTCTTGTAGACACACCTTTGCTGAAGATTCCT	2253
QY	721	ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu	740
Db	2254	ACTCAGATCAAGAAATCCAGAGTGTGGAATCTGGAGGGGAGAGTGTGCAAGAGCTC	2313
QY	741	SerValGluGluGlnIleLysArgAsnArgTrpTrpAspGluAspGluAspGluGlu	759
Db	2314	TCTGTGGAAGAACAGATAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG	2370
RESULT 4			
AK000372			
LOCUS	AK000372	2536 bp	mRNA
DEFINITION	Homo sapiens cDNA FLJ20365 fis, clone HEP17877.		linear
ACCESSION	AK000372		
VERSION	AK000372.1	GI:7020415	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (sites)		
	Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,		
	Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Ota, T.,		
	Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,		
	Nakamura, Y., Isogai, T. and Sugano, S.		
	NEDO human cDNA sequencing project		
TITLE	Unpublished		
JOURNAL	2 (bases 1 to 2536)		
REFERENCE	Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,		
AUTHORS	Shibahara, T., Tanaka, T. and Nakamura, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,		

University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)			
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).			
Location/Qualifiers			
1..2536	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="HEP17877"		
	/cell_line="HepG2"		
	/cell_type="hepatoma"		
	/clone_lib="HEP"		
	/note="cloning vector pME18SFL3"		
85..2199	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="BAA91120.1"		
	/db_xref="GI:7020416"		
	/translation="MESSPPNRRQMTSLSLRVTAKELSLVNKNKSSAIVEIFSKYOKA		
	ABETNMKXRSNTENLSQHRKGTITLVKKWENPCLGAESHTDSLRNSSTIRRAD		
	HPAEVTHSAAGAKADQEQIHPSRLASPPALVQGRYPHIKGEDLKHSTSSK		
	MENCGESHEVEKSEISENTDASGKIERIVFLNRLKMMPEKGTPTKILRAOSRS		
	ASGRKISENSYSLDLEIGPGQSSSTFDSKXNRNLELPLSETISKDMARKYQA		
	AVSKSSNTYTNELKASGEIKIHKMEQENVPVPEVCITHQEKISANENSLAV		
	RSTPAEDSDRSQVKEVQVHPKPLSPDSRASSLSSESSPPKAKKFCOPACTVE		
	COKTVYPMERLLANQOVFHSICFRCYCNKLISLGTYSALSHGRVYCKPHNQLFKSG		
	NYDEGFGHRPHKLMASKNENEILLERPAQLANARETPSPGVEDAPIAKVGLAASM		
	EKAKSSQKEDKPAETKKLRIAMPPPTTELSSGSGSALBEGIKMSKPPWPEDEISKPE		
	VPEVDLDLKLRRSSLSLKRSPFTVAASFQTSVSKPTKSPTRKQWSESESEE		
	SVGGRVAERKQVENAKSKNGNVKTTWQNKESKETGKRSKEGSHLEMNENLVEN		
	GADSDDDNSFLKHLKHNPSU		
BASE COUNT	863 a 551 c 612 g 510 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	7,42e-205	Length:	2536
-core:	3899.00	Matches:	758
Percent Similarity:	99.87%	Conservative:	0
Best Local Similarity:	99.87%	Mismatches:	1
Query Match:	99.29%	Indels:	1
DB:	9	Gaps:	0
US-09-890-549-4 (1-759) x AK000372 (1-2536)			
QY	1	MetGluSerSerProPheAsnArgArgGlnTrpThrSerLeuSerLeuArgValThrAla	20
Db	85	ATGGAATCATCTCCATTTAATAGCGCAATGACCTCACTATCATTTAGGGGTAAACGCC	144
QY	21	LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys	40
Db	145	AAAGAACCTTCTCTTGTCAACAAGAACCAAGTCATCGGCTATTGTGGAAATATTCTCAAG	204
QY	41	TyrGlnLysAlaAlaGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu	60
Db	205	TACCAGAAAGCAGCTGAGAAACAACATGGAGAGAGAGAGAAATACCCAAATCTC	264
QY	61	SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly	80
Db	265	TCCAGACATTTAGAAAGGGGACCTGACTGTGTTAAGAGAGAGTGGAGAACCCAGGG	324
QY	81	LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg	100
Db	325	CTGGAGCAGAGAGTCTCACACAGACTCTCTACGAAACAGCAGCAGCTGAGATTAGGCACAGA	384
QY	101	AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln	120

385 GCAGACCATCTCTGCTGAAGTGACAGCCAGCTGCTTCTGGAGCCAAAGCTGACCAA 444
QY
121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db
445 GAAGAACAAATCCACCCAGATCTAGACTCAGTCACTCCCTGAAAGCCCTCGTTGAGG 504
QY
141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db
505 CGATATCCCAACATCAAGAGCCGTGAGGATCTTAAAGACCATCTCAACAGAAAGTAA 564
QY
161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluLysSerGluAsn 180
Db
565 ATGGAATAATGCTAGGAGAAATCCAGGCATGAAGTAGAATAATCAGAAATCAGTGAA 624
QY
181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db
625 ACAGATGCTCTGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAGATG 684
QY
201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
Db
685 TTTGAGAAAGGTGAACCACTCAACTAAGATTCTCCGGGCCAAAGCCGAAAGTCAAGT 744
QY
221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
Db
745 GGAAGGAAGATCTCTGAAACAGCTATTCTCTAGATGACCTGGAATAGCCCCAGGTCA 804
QY
241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgAsnLeuGluLeuPro 260
Db
805 TTGTCACTCTTACATTTGACTCGGAGAAATAGAGTAGACGAAATCTGGAACCTTCCA 864
QY
261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db
865 CGCCTCTCAGAACCTCTCTATAAGAGTGAATGGCCAAAGTACGAGCAGCTGTGCCAA 924
QY
281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
Db
925 CAAGCAGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAAT 984
QY
301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db
985 CATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCTGTGAGTCTGCATCACCCTCA 1044
QY
321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db
1045 GAAGGGGAAAGATTTCTGCAAATGAGATAGCTGGCAGTCCGTTCCACCCCTGCCGAA 1104
QY
341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
Db
1105 GATGACTCCCGTGACTCCAGGTTAAGATGAGGTTCAACAGCCTGTCCATCCCAGCCA 1164
QY
361 LeuSerProAspSerArgAlaSerLeuSerGluSerSerProProLysAlaMetLys 380
Db
1165 CTAAGTCCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTTCTCTCCCAACCAATGAAG 1224
QY
381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
Db
1225 AAGTTTCAGGCACCTGCAAGAGACCTCGGTGGAATGTGAGAAGACAGTCTATCCAATG 1284
QY
401 GluArgLeuLeuAlaAsnGlnGlnPheHisIleSerCysPheArgCysSerTyrCys 420
Db
1285 GAGCGTCTTTGGCCCAACAGAGGTTTTCATCATCAGCTGCTTCGTTGCTCTTATTGC 1344
QY
421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
Db
1345 AACCAACAACTAGTCTAGAACATATGATCTTTACATGGAGAAATCTATTGTAAGCCT 1404
QY
441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db
1405 CACTTCAATCAACTTTTAAATCTAAGGCAACTATGATGAAGGCTTTGGGCACAGACCA 1464
QY
461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAlaGln 480
Db
1465 CACAAGGATCTATGGGCAAGCAAAATGAAACGAAAGAGATTTTGGAGAGACAGCCAG 1524

QY 481 LeuAlaPheAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1525 CTTCGAAATGCAAGGGAGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTTATTGCTAAG 1584
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
Db 1585 GTGGGTCTCTGGTGCAGATATGGAAGCCAGGCTCTCTCTCAGCAGGAGAAAGAGAC 1644
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProPheThrGluLeuGlySer 540
Db 1645 AAGCAGCTGAAACCAAGAAAGCTGAGGATCGCTGGCCACCCCTGAACTTGGAAAT 1704
QY 541 SerGlySerAlaLeuGluGluGlyLysMetSerLysProLysTyrProProGluAsp 560
Db 1705 TCAGGAAGTGCCTTGGAGGAAGGATCAAAATGTCAAAGCCCAATGCCCTCTCTGAAAGAC 1764
QY 561 GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
Db 1765 GAATTCAGCAAGCCCGAAGTTCTGAGATGTCATCTAGATCTGAAAGACCTAAGACGA 1824
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
Db 1825 TCTTCTTCTACAGGAAAGAGCGGCCCTTACTGTAGCAGCTTCTATTCAAGCACC 1884
QY 601 SerValLysSerProLysThrValSerProProIleArgLysGlyTyrSerMetSerGlu 620
Db 1885 TCTGTCAAGACCCCAAAACTGTGTCCCACTATCAGAAAGGCTGGAGCATGTCCAGAG 1944
QY 621 GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
Db 1945 CAGAGTGAAGAGTCTGTGGGTGGAAGATGTCAGAAAGGAAACAAAGTGGAAATGCCAAG 2004
QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
Db 2005 GCTTCTAAGAGAAATGGGAATGCGGAAACAAACCTGGCNAACAAAGAAATCTAAAGGA 2064
QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 2065 GAGCAGGGAAGAGAGATGAGAAAGTCTAGTTTGGAGATGGAGATGAGATCTTGTA 2124
QY 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGln 700
Db 2125 GAAATGGTCAGACTCCGATGAAGATGATAACAGCTTCTCTCAACCAACA-TCTCCACA 2183
QY 701 GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPheThr 720
Db 2184 GAACCCAAAGTCTCGAATTTGGTCGAGTTTGTAGACCAACACCTTTGTCTGAAAGAAATCCACT 2213
QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu 740
Db 2244 ACTCAGATATCAGAAATCCAGGATGTGAACCTCTGGAGGGAGAGTGGTCAAAGAGCTC 2303
QY 741 SerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 2304 TCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2360

RESULT 5

AF198455

LOCUS

DEFINITION Homo sapiens epithelial protein lost in neoplasm alpha (EPLIN)-
mRNA, complete cds.

ACCESSION AF198455

VERSION AF198455.1

GI:6685008

SOURCE

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3550)

Maul, R.S. and Chang, D.D.

EPLIN, epithelial protein lost in neoplasm

Oncogene 18 (54), 7838-7841 (1999)

MEDLINE 20087188
 PUBMED 10618726
 REFERENCE 2 (bases 1 to 3550)
 AUTHORS Maul,R.S. and Chang,D.D.
 TITLE Direct Submission
 JOURNAL Submitted (25-OCT-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA
 FEATURES
 source Location/Qualifiers
 1. 3550
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="12"
 /map="12q13"
 1. 3550
 /gene="EPLIN"
 474..2276
 /gene="EPLIN"
 /codon_start=1
 /product="epithelial protein lost in neoplasm alpha"
 /protein_id="AAF23756.1"
 /db_xref="GI:6685009"
 /translation="MENCLESREHEVKSISENTDASGKIEKNVPLNELKVMFSK
 EPTQTILRAQSASGRKISSENSYDLDLEIGPQLSSSTFDSEKNESRNLELPR
 SEISIKDRMAKYQAASVKQSSNITNELLKASGIEKHKMEQKENVPPGEVCITHT
 AMKISANESLAVRSTPAEDDSPGDSQVSEVQPPVHPKPLSPDGRASSLSPPK
 EGKIFQAPARETCEVCQTPYPMERLLANQQVFHISCFRSCYCNKLISLGTYSALHGR
 IYCKPHNQPLFKSGNYDEGFGPHRPHKDLWASKNEEILERPAAQLANARETPHSPGV
 EDAPIAKGVLAASMEAKASQOEKDKPAETKKLRIAMPPPTLELSSGSALSEGIVN
 SKYKWPEDISIPVEDVDLKLKLRSSSLKERSRPETVAASQSTSVK3PKTVS
 PPRIKMSMEQEEBESVGGVARERKQVAKSKNGNNGVKTWNKESKGTGKRSK
 EGHSLMENLVLVNGADSDDDNSFLKQSQPOEPKSLNWSFVDNTFAEEFTQNK
 SQDVLWEVGEVVKLSVEEQIKRNYDEDEDE"
 3527..3532
 polyA_signal
 BASE COUNT 1153 a 716 c 773 g 908 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,61e-190 Length: 3550
 Score: 3640.50 Matches: 703
 Percent Similarity: 99.72% Conservative: 0
 Best Local Similarity: 99.72% Mismatches: 1
 Query Match: 92.70% Indels: 1
 DB: 9 Gaps: 1
 US-09-890-549-4 (1-759) x AF198455 (1-3550)
 Qy 56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys 75
 Db 159 AACACCGAAATATCTCTCCAGCACTTTAGAAAAGGGGACCCCTGACTGTGTTAAAGAAGAAG 218
 Qy 76 TrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThr 95
 Db 219 TGGGAGAACCCAGGCGTGGGAGCAGAGTCTCACACAGACTCTCTACGGAAACAGCAGCACT 278
 Qy 96 GluIleArgHisArgAlaAspHisProAlaGluValThrSerHisAlaAlaSerGly 115
 Db 279 GAGATTAGGCACAGACGAGACCATCTCTCTGCTGCTGAAGTGCACGCCACGCTGCTTCTTGG 338
 Qy 116 AlaLysAlaAspGlnGluGluGlnIleHisProArgSerArgLeuArgSerProProGlu 135
 Db 339 GCCAAAGCTGACCAAGAAGAACAATCCACCCAGATCTAGACTCAGCTCACCTCTCTGAA 398
 Qy 136 AlaLeuValGlnGlyArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSer 155
 Db 399 GCCCTCGTTGAGGGTTCGATATCCCCATCAAGACGGTGAGGATCTTAAAGACCACTCA 458
 Qy 156 ThrGluSerLysLysMetGluAsnCysLeuGlyLysSerArgHisGluValGluLysSer 175
 Db 459 ACAGAAAGTAAAAAATGGAAAAATTGCTAGGAGAAATCCAGGCATGAAGTAGAAAAATCA 518
 Qy 176 GluIleSerGluAsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsn 195

```
Db 1599 CCCACTGAATCTTGGAAAGTTTCAGGAAGTGGCTTGGAGGAAGGATCAAAAATGTCAAAGCCC 1658
Qy 555 LysTrpProGluAspGluLeuSerLysProGluValProGluAspValAspLeuAsp 574
Db 1659 AAATGGCCCTCCTGAAGACCAATCAGCAAGCCCGAAGTTCCTGAGGATGCTGATCTAGAT 1718
Qy 575 LeuLysLysLeuArgSerSerLeuLysGluArgSerArgProPheTrnValAla 594
Db 1719 CTGAAGAACTAAGACGATCTTCTTCACTGAAGAAAGGCGCCCATCTACTGTAGA 1778
Qy 595 AlaSerPheGlnSerThrSerValLysSerProLysThrValSerProProleArgLys 614
Db 1779 GCATCAATTTCAAAGCACCTCTGTCAAGAGCCCAAAAACCTGTGCCACCTATCAGGAAA 1838
Qy 615 GlyTrpSerMetSerGluGlnSerGluGluSerValGlyArgValAlaGluArgLys 634
Db 1839 GCGTGGAGCATGTCAGAGCAGAGTGAAGAACTCTGTGGTGGAGAGTTCAGAAAAGGAAA 1898
Qy 635 GlnValGluAsnAlaLysLysLysLysAsnGlyAsnValGlyLysThrThrTrpGln 654
Db 1899 CAAGTGGAAAATGCCAAGGCTTCTAGAAAGATGGGAAATGGGAAAACCAACCTGGGAAA 1958
Qy 655 AsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMet 674
Db 1959 AACAAAGAAATCTAAAGGAGACAGGGAAGAGTAAGGAAGTCTATAGTTGGAGATG 2018
Qy 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeu 694
Db 2019 GAGAATGAGAATCTTGTAGAAAATGCTGCAGATCCGATCAAGATGATTAACAGCTTCTCT 2078
Qy 695 LysGlnGlnSerProGlnProLysSerLeuAsnTrpSerSerPheValAspAsnThr 714
Db 2079 AACAAACAATCTCCAAAGAACCCAAAGTCTCTGAATGGTGGAGTTTGTAGAACACAC 2138
Qy 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly 734
Db 2139 TTTGCTGAAGAAATCTACTACTAGATCAGAAATCCAGAGATGGAATCTTGGAGGGA 2198
Qy 735 GluValValLysGluLeuSerValGluGluGlnLeuLysArgAsnArgTrpTrpAspGlu 754
Db 2199 GAAGTGGTCAAGAGCTCTGTGGAGAACACAGATTAAGAGAAATCGGTATTATGATGAG 2258
Qy 755 AspGluAspGluGlu 759
Db 2259 GATGAGGATGAAGAG 2273

RESULT 6
BC001247 2164 bp mRNA linear PRI 12-JUL-2001
LOCUS
DEFINITION
Homo sapiens, epithelial protein lost in neoplasm beta, clone
MGC:4969 IMAGE:3452714, mRNA, complete cds.
ACCESSION
BC001247
VERSION
BC001247.1 GI:12654808
KEYWORDS
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2164)
Strausberg,R.
Direct Submission
Submitted (11-DEC-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
```

Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: JRAK Plate: 4 Row: 0 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7020350.
Location/Qualifiers
1. 2164

FEATURES

source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:51474"
/db_xref="taxon:9606"
/clones="MGC:4969 IMAGE:3452714"
/tissue type="placenta, choriocarcinoma"
/clone_lib="NIH_MGC_10"
/lab_host="DH10B"
/note="vector: pCMV-Sport6"
187..1983
/codon_start=1
/product="epithelial protein lost in neoplasm beta"
/protein_id="AAH01247.1"
/db_xref="GI:12654809"

CDS

BASE COUNT 744 a 472 c 513 g 435 t
ORIGIN
Alignment Scores:
Pred. No.: 1,49e-178 Length: 2164
Score: 3417.50 Matches: 660
Percent Similarity: 99.85% Conservative: 0
Best Local Similarity: 99.85% Mismatches: 0
Query Match: 87.03% Indels: 1
DB: 9 Gaps: 1
US-09-890-549-4 (1-759) x BC001247 (1-2164)
Qy 99 HisArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAla 118
Db 1 CACAGAGCAGACCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
Qy 119 AspGlnGluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuVal 138
Db 61 GACCAAGAGACAAATCCACCCAGATCTAGATCTAGATCTAGATCTAGATCTAGATCTAGAT 120
Qy 139 GlnGlyArgTrpProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSer 158
Db 121 CAGGTCGATATCCACATCAAGACCGTGAGGATCTTAAAGACCATCTCAACAGAAAGT 180
Qy 159 LysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluLysSer 178
Db 181 AAAAAATGGAAAAATTTGTAGGAGAAATCCAGGATCAAGTAAAGTAAAGTAAAGTAAAGT 240
Qy 179 GluAsnThrAspAlaSerGlyLysIleGluLysTrpAsnValProLeuAsnArgLeuLys 198
Db 241 GAAAAACAGATGCTTCGGGCAAAATAGAGAAATATATATGTTCCGCTGAACAGGCTTAAG 300
Qy 199 MetMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSer 218
Db 301 ATGATGTTTGAGAAAGGTGAACCAACTCAACTCAACTCAACTCAACTCAACTCAACTCAACT 360

QY 219 AlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyPro 238
 DB 361 GCAAGTGGAGGAAGATCTCTGAAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCCA 420
 QY 239 GlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgAsnLeuGlu 258
 DB 421 GGTCAATTGTCATCTTCTACATTTGACTCGGAGAAAATGAGAGTAGACGAAATCTGGAA 480
 QY 259 LeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaVal 278
 DB 481 CTTCCAGCCCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCAGGCGCTGTG 540
 QY 279 SerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGluIle 298
 DB 541 TCCAAACAAACAGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGTGGCGAAATC 600
 QY 299 LysIleHisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThr 318
 DB 601 AAAATTCTAATAATGGACCAAAAGAGATGTGCCCCAGGTCTGAGGTCTGCATCACC 660
 QY 319 HisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrPro 338
 DB 661 CATCAGGAAGGGAAAAGATTCTTGCAATGAGAATAGCTGGCAGTCCGTTCCACCCCT 720
 QY 339 AlaGluAspAspSerArgAspSerGlnValLysSerGluValGlnProValHisPro 358
 DB 721 GCCGAAGATGACTCCCGTGACTCCAGGCTTAAAGAGTGAAGGTTCAACAGCCGTGCCATCCC 780
 QY 359 LysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAla 378
 DB 781 AAGCCACTAAGTCCAGATCCAGAGCCTCCAGTCTTTCTGAAAGTCTCTCCCAAGCA 840
 QY 379 MetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyr 398
 DB 841 ATG---AAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGATGTCAGAGACAGTCTAT 897
 QY 399 ProMetGluArgLeuLeuAlaAsnGlnValPheHisLysSerCysPheArgCysSer 418
 DB 898 CCAATGGAGCGTCTCTTGGCCAAACAGCAGGTGTTTCACATCAGCTGCTCCGTTGCTCC 957
 QY 419 TyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCys 438
 DB 958 TATTGCCAACCAACAACCTCAGTCTAGGAACATATGCATCTTTACATGGAAGAACTATTGT 1017
 QY 439 LysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHis 458
 DB 1018 AAGCTCACTTCAATCACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGCAC 1077
 QY 459 ArgProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluIleLeuGluArgPro 478
 DB 1078 AGACCACAAAGGATCTATGGCAAGCAAAAATGAAACGAAGAGATTTTGGAGAGACCA 1137
 QY 479 AlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIle 498
 DB 1138 GCCCAGCTTGAATGCAAGGGAGACCCCTCACAGCCCGAGGGTAGAAGATGCCCTATT 1197
 QY 499 AlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLys 518
 DB 1198 GCTAAGTGGGTGTCCTGGCTGCAAGTATGGAACCAAGCCCTCTCTCAGCAGGAGAAG 1257
 QY 519 GluAspLysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeu 538
 DB 1258 GAACACAAGCCAGCTGAAACCAAGACCTGAGGATCGCTGGCCACCCCTCACTGACTT 1317
 QY 539 GlySerSerGlySerAlaLeuGluGluGlyLysMetSerLysProLysTrpProPro 558
 DB 1318 GGAAGTTTCAGGAGTGGCTTGGAGAGGAGTCAAAATGTCAAAGCCCAATGGCCTCT 1377
 QY 559 GluAspGluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLeu 578
 DB 1378 GAAGCGAAATCAGCAAGCCGAAAGTCTCTGAGGATGTCATCTAGATCTGAAAGAGCTA 1437

QY 579 ArgArgSerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGln 598
 DB 1438 AGACGATCTTCTTCACTAGAGAAAGAGCCGCCCATTCCTGTAGCAGCTTCATTTCAA 1497
 QY 599 SerThrSerValLysSerProLysThrValSerProProIleArgLysGlyTrpSerMet 618
 DB 1498 AGCAGCTCTGTCAAGAGCCCAAAACTGTGTCCCACTATCAGAAAGGCTGGAGCATG 1557
 QY 619 SerGluGlnSerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsn 638
 DB 1558 TCAGAGCAGAGTGAAGAGTCTGTGGTGGAGAGTTCAGAAAGAAACAAGTGGAAAT 1617
 QY 639 AlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSer 658
 DB 1618 GCCAAGGCTTCTAAGAAGAAATGGAAATGTGGGAAAAACAACCTGGCAAAACAAGAATCT 1677
 QY 659 LysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetClnAsnGluAsn 678
 DB 1678 AAAGAGAGACAGGAGAGAGAGTAAAGAGGTCTAGTTTGGAGATGGAGATGAGAAAT 1737
 QY 679 LeuValGluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLysGlnGlnSer 698
 DB 1738 CTTGTAGAAAATGGTGCAGACTCCGATGAGATGATACAGCTTCTCAACAACAATCT 1797
 QY 699 ProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGlu 718
 DB 1798 CCACAAGAACCCCAAGTCTCTGAATTTGGTGGAGTTTGTAGACAACACCTTTCTCTGAAGAA 1857
 QY 719 PheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValLys 738
 DB 1858 TTCACTACTCAGAAATCCGAAATCCCAAGGATGGAACTCTGGAGGGAGAACTGGTCAAA 1917
 QY 739 GluLeuSerValGluGluGlnLysLysArgAsnArgTyrTyrAspGluAspGluAspGlu 758
 DB 1918 GAGCTCTCTGTGGAGACACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAA 1977
 QY 759 Glu 759
 DB 1978 GAG 1980

RESULT 7
 AK000335 3277 bp mRNA linear PRI 22-FEB-2000
 LOCUS Homo sapiens cDNA FLJ20328 fis, clone HEP10039.
 DEFINITION
 ACCESSION AK000335
 VERSION AK000335.1 GI:7020350
 KEYWORDS cDNA capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
 Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
 Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
 Nakamura,Y., Isogai,T. and Sugano,S.
 NEDO human cDNA sequencing project
 Unpublished
 1 (bases 1 to 3277)
 2 (bases 1 to 3277)
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
 Shibahara,T., Tanaka,T. and Nakamura,Y.
 Direct Submission
 Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
 Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
 Tel:81-3-5449-5286, Fax:81-3-5449-5416)
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing; Research Association for Biotechnology; cDNA library
 construction, 5'- & 3'-end one pass sequencing; Department of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).

FEATURES		Location/Qualifiers	
source		1..3277	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="HEP10039"	
		/cell_line="HepG2"	
		/cell_type="hepatoma"	
		/clone_lib="HEP"	
		/notes="Cloning vector pME18SFL3"	
CDS		179..1978	
		/notes="unnamed protein product"	
		/codon_start=1	
		/protein_id="BAA91092.1"	
		/db_xref="GI:702031"	
		/translation="MENCLEGRHEVEKSEISNTDASOKIEKYNVPLNRLRMPEKG EPTQKILRAQSRGKRSISNSYSLDLEIGPQLSSSTFDSEKNESRRNLEIPRL SETSIDRMWAKTQAASVKSSTNYTNELKASGEIKHKMEQENVPVPGVCVITHQ EKEKISANESLAVRTPDDSDRSQVSEVQOPVHPKPLSPDRASLSSESPPKA MKKFOAPARETCVCKQTVYPMERLLANQOVPHISCFRCSCYNKSLGTYASLHGR YCKPHNFQKSKNGTDEGHRPHDKLWASKNENEILLERPAQLANARETHPSGVE DAPIAKGVLAASMEKASQOEKDPKPAETKXLRAMPPTTELGSGLSALBEGIKMS KPKPPDEISKPEVEDVDLKLRRSSSLKRSRRPPTVAASQSTSVKSPKTVSP PIRKGMSEQSESVGGVABRKOVENAKSKXGNVGTWONKESKGETGRSKE GHSLENNENLVENGADSDSDNSFLKQSPQEPKSLNWSFVDNFAEEFTTQKRS QDVELWEGEVVLSVEEQIKRNRYDEDEDEE"	
BASE COUNT		1089 a 651 c 703 g 834 t	
ORIGIN		179..1978	
Alignment Scores:		2.94e-178 Length: 3277	
Pred. No.:		3416.00 Matches: 658	
Score:		100.00% Conservative: 0	
Percent Similarity:		100.00% Mismatches: 0	
Best Local Similarity:		86.99% Indels: 0	
Query Match:		9 Gaps: 0	
DB:		US-09-890-549-4 (1-759) x AK000335 (1-3277)	
QY	102	AspHisProAlaGluValThrSerHisAlaAlaSerGlyValAlaLysAlaAspGlnGlu	121
DB	2	GACCAATCTCTGCTGAAGTGACAGCCAGCTGCTTCTGGAGCCAAAGCTGACCAAGAA	61
QY	122	GluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGlyArg	141
DB	62	GAACAAATCCACCCAGATCTAGACTCAGCTCACCTCTCTGAAAGCCCTCGTTCCAGGTCGA	121
QY	142	TyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLysMet	161
DB	122	TATCCCCACATCAAGACCGGTGAGGATCTTAAAGACCACCTCAACAGAAATGAAAAATG	181
QY	162	GluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsnThr	181
DB	182	GAATAATTGCTAGGAGATCCAGGATGAGTAGAATAATCAGAATCAGTGAATAACACA	241
QY	182	AspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMetPhe	201
DB	242	GATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGTCGACAGGCTTAAGATGATGTT	301
QY	202	GluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSerGly	221
DB	302	GAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGGCCCAAGCCGAAGTCAAGTGA	361
QY	222	ArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGlnLeu	241
DB	362	AGAGAGATCTCTGAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCAGGTGAGTTG	421
QY	242	SerSerSerThrPheAspSerGluLysAsnGlnSerArgArgAsnLeuGluLeuProArg	261
DB	422	TCATCTCTTACATTTGACTCGGAGAAAAATGAGATGACGCAAAATCTGGAACTTCCACGC	481
QY	262	LeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLysGln	281
DB			
DB	482	CTCTCAGAAACCTCTATATAAGATCGAATGCCAAGTACCAGGAGCTGTGTCCAAACAA	541
QY	282	SerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluLysIleHis	301
DB	542	AGCAGCTCAACCAACTATACAAATGAGCTGAAGACGAGTGGTGGCGAAATCAAAATTCAT	601
QY	302	LysMetGluGlnLysGluAsnValProProGlyLysProGluValCysIleThrHisGlnGlu	321
DB	602	AAATGGGAGCAAAAGAGAAATGTGCCCCCAGGTCTCTGAGGTCTGCATCACCCTCAGGAA	661
QY	322	GlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGluAsp	341
DB	662	GGGGAAGAGATTTCTGCAATAGCAATAGCTGGCAGTCCGTTCCACCCCTGCCGAGAT	721
QY	342	AspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysProLeu	361
DB	722	GACTCCCGTACTCCAGGTTAAGAGTGAGGTCAACAGCCGTGTCCATCCCAAGCCACTA	781
QY	362	SerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLysLys	381
DB	782	AGTCAGATTCCAGAGCCTCCAGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAG	841
QY	382	PheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMetGlu	401
DB	842	TTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTGAGAAAGACAGTCTATCCCAATGGAG	901
QY	402	ArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCysAsn	421
DB	902	CGTCTCTGGCCCAACAGCAGCTGTTTACATCAGCTGCTTCCGTGCTCTCTATTGCAAC	961
QY	422	AsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysProHis	441
DB	962	AACAAACTCAGCTAGGAACATATGCTCTTTTACATGGAAGAACTTATTGTAAGCCTCAC	1021
QY	442	PheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgProHis	461
DB	1022	TTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACAC	1081
QY	462	LysAspLeuTyrAlaSerLysAsnGluGluLeuLeuGluArgProAlaGlnLeu	481
DB	1082	AAGATCTATGGGCAAGCAAAATGAAAGCAAGAGATTTTGGAGAGACCAGCCAGCTT	1141
QY	482	AlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLysVal	501
DB	1142	GCAATGCAGGGAGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTTATTGCTAAGGTG	1201
QY	502	GlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAspLys	521
DB	1202	GGTGTCTGCTGCTCAAGTATGGAAGCCAGCCCTCTCTCAGCAGGAGAGAGAGACAAG	1261
QY	522	ProAlaGluThrLysLysLeuArgIleAlaTyrProProProThrGluLeuGlySerSer	541
DB	1262	CCAGTGAACCAAGAGAGCTGAGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTCA	1321
QY	542	GlySerAlaLeuGluGlyIleLysMetSerLysProLysTyrProProGluAspGlu	561
DB	1322	GGAAGTCCCTTGGAGGAAGGATCAAAATGTCAAGCCCAAAATGGCTCTCTGGAAGCGAA	1381
QY	562	IleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArgSer	581
DB	1382	ATCAGCAAGCCCAAGTTCTCTGAGGATGTGATCTAGATCTGAGAGAGCTTAAGACGATCT	1441
QY	582	SerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThrSer	601
DB	1442	TCCTTCACTGAAGGAAGAGAGCCGCCCTTCACTGTAGCAGCTTCACTTCAAGACCTCT	1501
QY	602	ValLysSerProLysThrValSerProProIleArgLysGlyTyrSerMetSerGluGln	621
DB	1502	GTCAAGAGCCCAAAACTGTGTCTCCACCTATCAGGAAGAGCTGGAGCATGTGAGAGCAG	1561
QY	622	SerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLysAla	641
DB	1562	AGTGAGAGATCTGTGGGTGGAGAGATTGCAAGAAAGGAACAGTGGAAATGCCAAGGCT	1621

642 SerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGlyGlu 661
Db TCTAAGAAGAAATGGGAATGTGGGAAACCAACCTGGCAAAACCAAGAATCTAAGAGAGAG 1681
662 ThrGlyLysArgSerLysGlyGlyHisSerLeuGluMetGluAsnGluAsnLeuValGlu 681
Db ACAGGAGAGAGAGTAAGAGAGTCAAGTTTGGAGATGGAGATGAGAAATCTTGGAA 1741
682 AsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGlnGlu 701
Db AATGGTCAGACTCCGATGAAGATGATGAACAGCTTCTCAACACCAATCTCCACAGAA 1801
702 ProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheLaGluGluPheThrThr 721
Db CCCAAGTCTCTGAATGGTGCAGTTTGTAGACAAACACCTTGTGTAAGAATCTCACTACT 1861
722 GlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeuSer 741
Db CAGAAATCAGAAATCCAGGATGTGGAATCTTGGAGGGGAGAGTGGTCAAGAGCTCTCT 1921
742 ValGluGluGlnLysLysArgAsnArgTyArgAspGluAspGluGluGlu 759
Db GTGGAAGACAGATAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 1975

RESULT 8
BC031490 3232 bp mRNA linear ROD 20-SEP-2002
LOCUS Mus musculus, epithelial protein lost in neoplasm, clone MGC:27894
DEFINITION IMAGE:3498139, mRNA, complete cds.
ACCESSION BC031490
VERSION BC031490.1 GI:21619380
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3232)
Strausberg, R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: capbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 35 Row: 1 Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 12746445.
Location/Qualifiers
1. 3232
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:27894 IMAGE:3498138"
/tissue_types="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."

/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
81. 2342
/codon_start=1
/product="epithelial protein lost in neoplasm"
/protein_id="AAH31490.1"
/db_xref="GI:21619381"
/db_xref="LocusID:65970"
/translation="WESTPFRNRROWTSLSRLVTKELSLVKNKSSAIVEIFSKYOKA
AEEANMSEKKNPESLPQHFRRGTLVLKXKWPVAGAEPTDSLPSSEGGHTAD
YPPAEVTPKPGVADREHTQKPRFGSRPEAVIQSRYPRSENHDPKQAATESQK
MENCUGDRHEAEKPESENTETSGKIEKYNPLNRLKMFKEGHEHNTKSLMTQSRN
AGGRLSNCSLDDWEIGAGLSNNAFSEKSNKLELPRISETIKORMARYQA
AVSKOSSPASYNELKTSKTHKWEQENVPGPPEACSVHQEGSKVSTTNSLVALS
VPAEDTCNSOVKSEAQQPHKPLSPDARTSSLPSSPSKTKKFOAPAKESCVECO
KTVPMESLLANQOVFHSICRSCYCNKLSLGTYSALHGRILYCKPHRNLQPKSGNY
DEGFOHKQKOLWASKSDNEETLGRPAQPPNAGESPHSPGVEDAPIAKGVLAASMEA
KASSOREDERPAETKLRILIAWPPPAELGSGSALLEGIKVSKPKWPPEDDVCKTEAP
EDVDLKLKRRSLKERSRPFVAASFRSTSIKSPKASPSLRKMSSEGESEBEFG
GGIATMERKOTENARPSCGEKENVGSRWQGEVPRSKDRSSFELESENFMEGANGIAE
DDNHVHAQOSPLEPEAPQWGSFVDTTAAKEFTTQNKRSQDVGFWEVEVRLSVEEQI
KRNRYDEDEEE"

BASE COUNT 969 a 792 c 853 g 618 t
ORIGIN
Alignment Scores:
Pred. No.: 7,936-154 Length: 3232
Score: 2970.00 Matches: 577
Percent Similarity: 83.82% Conservative: 60
Best Local Similarity: 75.92% Mismatches: 115
Query Match: 75.63% Indels: 8
DB: 10 Gaps: 3

US-09-890-549-4 (1-759) x BC031490 (1-3232)
QY 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db ATGGAAATCGACTCCATTTAATAGACCCAGTGGACTTCCCTGTCTATTGAGAGTAAACGCC 140
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db AAGAGAGCTTCTCTTGTCAACAGAAACAAGTCATCCGCAATTTGTGAAATATTCTCCAAG 200
QY 41 TyrGlnLysAlaAlaGluGluThrAsnValLysLysLysArgSerAsnThrGluAsnLeu 60
Db TACCAGAGAGGCTGCTGAAGAAAGCCCAACATGGAAGAGGAGAAATAAACCCTCAAGTCTG 260
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db CCCAGCAGCTTTTAGAAGAGGAGCCCTGCTGTGTTAAAGAAAGAGTGGGAAACCCCGTG 320
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db GCTGGGGCAGAAATTCACACAGACTCACTGCCAACACAGCAGCAGTGGGGTGGGCACACA 380
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
Db GCGGACTACCTCTGCTGAAGTACGCGACCAAGCCTGCTCTCTGGAGTCAAGTGGAGTGG 440
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db GAAGAGCACACCCAGCCCAACCTAGATTTCGATCTCTCCGGAAGCAGTATTCCAGAGC 500
QY 141 ArgTyrProHisIleLysAspGlyGluAspLysAspHisSerThrGluSerLysLys 160
Db CGGTATCTCTCGCTCAGAGAACAGCCACGATTTTAAAGCCAGCCAGGAGGCCAATAA 560
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGlnLysSerGluIleSerGluAsn 180
Db ATGGAATACTGCTGGGAGATTCAGGAGCATCAACAGAGAGCCAGAGAGCCAGAGCGAANAAC 620
QY 181 ThrAspAlaSerGlyLysIleGluLysTyArgAsnValProLeuAsnArgLeuLysMetMet 200

Db 621 ACAGAACTTCGCGCAAAATAGAGAAATACAGTTCCTCACTGAATAGATGAAGATG 680
QY 101 PheGluLysGlyGluProThrGlnThrLysIleLeuAArgAlaGlnSerArgSerAlaSer 220
Db 681 TTTGAGAAAGGTGAACACAAACCAACCAAGAGTCTCTGGAGCCCAAGCCCAAAATGCGGT 740
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
Db 741 GGAAGAGAGGTCTCTGAAACAACTGTTCCCTGGATGACTGGGAAATAGTGGCGGTCA 800
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db 801 TTCTCTCTCTCTCACTTCGAGTGGAGAAACAGAGAGTAAGAGGAAATCTGAGCTGCCA 860
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db 861 CGCCTCTCGAAACCTCCATAAAGGACCGCATGGCCAAAGTACCAGGCTCAGTGTCCAAG 920
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGluLysIle 300
Db 921 CAGAGAGCCCGCCAGCTACCAAAATGAGCTGAAACACAGT-----GAAAGCAAACT 974
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db 975 CATAATGGACAGAGAGAGATGTGCCCCAGGTCCCGAGGCTCGACGCTGCATCAG 1034
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 1035 GAAGGAAGCAAGTTCTCAACTGAGATAGTCTAGTGGCCCTCTCTGCTCCTGCTGAA 1094
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
Db 1095 GATGACACCTGTAACTCCAGGTGAAGAGCGAGGCCAGCGCTATGCAACCTTAAGCCG 1154
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1155 CTGAGTCTAGCCAGGACCTCCAGTCTCCGAAAGTCTCCTTCCAAAACACGCGAAG 1214
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
Db 1215 AAGTTTCAGCGCCGCAAGAGAGAGCTCGTGGAGTGTCAAGAACCGGTGTACCCCAATG 1274
QY 401 GluArgLeuLeuAlaAsnGlnValPheHisLeSerCysPheArgCysSerTyrCys 420
Db 1275 GAACGCTCTCTGCCAACAGCAGAGTGTTCACATCAGCTGTTCCGATGCTCTCTACTGC 1334
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
Db 1335 AACAAACAGCTCAGTCTAGGAACATATGATCCTTACATGGACGAATCTACTGTAAAGCT 1394
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1395 CACTTCAATCACTCTTCAAGTCCAAAGGCACTATGACGGGGTTTGGGCACAAAGCAG 1454
QY 461 HisLysAspLeuTyrPalaSerLysAsnGluGluIleLeuGluArgProAlaGln 480
Db 1455 CATAAGGATCTGTGGCAAGCAAGACCAATGAGGAGACTTTGGGAGACCCAGCCAG 1514
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1515 CGGCTAATGCGGGAGAGCCCAATAGCCGGGTGTAGAAGATGCCCCCATGCCAAG 1574
QY 501 ValGlyValLeuAlaAspMetGluAlaLysAlaSerSerGlnGlnLysGluAsp 520
Db 1575 GTCGCGTGTGGCGGCAAGTATGGAACCAAGGCCCTCTCTCAGAGGGAGAGGAGAT 1634
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTyrProProThrGluLeuGlySer 540
Db 1635 AAGCTCGGGAGACCAAGAGCTGAGGATCGCTGGCTCCGCCAGCGAGCTGGCGGT 1694
QY 541 SerGlySerAlaLeuGluGluGlyLysMetSerLysProLysTyrProProGluAsp 560

Db 1695 TCCGGAAGTGCCTGGAGGAAGGATCAAGATATCGAAGCCCAAGTGCCTCCGGAGGAT 1754
QY 561 GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
Db 1755 GACGCTCTCAAGACGAGCGCCCGGAGGATCTAGATCTGGACCTGAACAACTGCGGCG 1814
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
Db 1815 TCTTCGTCCTGAAAGAGAGGAGCGCCGCTTTACGCTAGCAGCTTCTTTTGAACCTCT 1874
QY 601 SerValLysSerProLysThrValSerProIleArgLysGlyTyrSerMetSerGlu 620
Db 1875 TCCATCAAGACCCCAAGGCCCTCTTGGCCATCTCTCAGAAAGGTGGAGCGAGTCCGAG 1934
QY 621 GlnSerGluGluSerValGlyGly---ArgValAlaGluArgLysGlnValGluAsnAla 639
Db 1935 CAGAGTGGAGGATTTGGAGGAGGATAGCAGCAGATGGAAGGAAACAAACGAAATGCC 1994
QY 640 LysLalaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLys 659
Db 1995 AGACCTCTGGGAGAAAGAGAATGTGGAAATCAACGCTGGCAG----- 2039
QY 660 GlyGluThrGlyLysArgSerLysGlyGlyHisSerLeuGluMetGluAsnGluAsnLeu 679
Db 2040 GCGGAGGAAGTCCCGAGGAGCAAGGACCGTAGTCTTTTGGAGCTGGAGAGTGAATTTT 2099
QY 680 ValGluAsnGlyAlaAspSerAspAspAspAsnSerPheLeuLysGlnSerPro 699
Db 2100 ATGGAATATGGAGCAACATAGCTGAAGATGACAAACCATGTCCACGACAGCTCCCG 2159
QY 700 GlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPhe 719
Db 2160 TTAGAACCCGAGGCTCCGGGTGGTCTGGCTTTGTACACACCACCGCTGCTAAAGATTC 2219
QY 720 ThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGlu 739
Db 2220 ACTACCCAGAAATCAAAATCCCAAGATGTCCGGTCTTGGGAGGAGAGTGGTCCGAGAG 2279
QY 740 LeuSerValGluGluGlnLysLeuLysArgAsnArgTyrTyrAspGluAspGluGlu 759
Db 2280 CTGTCTGTGGAGGAACAGATAAAGAGAAACCGGTACTATGATGAGGACGAGATGAAGA 2339
RESULT 9
AF307845 2370 bp mRNA linear ROD 02-MAR-2001
LOCUS Mus musculus epithelial protein lost in neoplasm-b (Eplin) mRNA,
DEFINITION complete cds.
ACCESSION AF307845
VERSION AF307845.1 GI:11127934
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2370)
AUTHORS Maul,R.S., Sachi Gerbin,C. and Chang,D.D.
TITLE Characterization of mouse epithelial protein lost in neoplasm
(EPLIN) and comparison of mammalian and zebrafish EPLIN
JOURNAL Gene 262 (1-2), 155-160 (2001)
MEDLINE 21100452
PUBMED 11179679
REFERENCE 2 (bases 1 to 2370)
AUTHORS Maul,R.S. and Chang,D.D.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2000) Medicine, UCLA, 10833 Le Conte Avenue, Los
Angeles, CA 90095, USA
FEATURES
source Location/Qualifiers
1..2370
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
1..2370
/gene="Eplin"

649	ACAGAAACTTTCAGGCAAAATAGAGAAATATACAACTGTTCCACTGAATAGACTGAAGATGATG	708
201	PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer	220
709	TTTGAGAAAGGTGAACACAAACCAACCAAGAGTCTCTGGACCCAAACCGAAATCGGGT	768
221	GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln	240
769	GGAGGAGGCTCTCTGAAACCAACTCTTCCTCGATGCTGGGAAATAGTTCGGGTTCAT	828
241	LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro	260
829	TTGTCTTCCTCTGCATTCAACTCGAGAAAAACGAGAGTAAGAGGAATCTGGAGCTGCCA	888
261	ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaIleValSerLys	280
889	CCCTCTCTGAAACCTTCATAAAGACCCGATGGCCAAAGTACGAGCTCGAGTGTCCAAG	948
281	GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle	300
949	CAGAGCAGCCAGCCAGCTACACAAATGAGCTGAAACCCAGT-----GAAACCAAACT	1002
301	HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln	320
1003	CATAAATGGGAACAGAGGAGAAATGTGCCCCAGGTCCGAGGCCTGCAGCGTCCATCAG	1062
321	GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu	340
1063	GAGGAGCAAGAGTTTCTCAACTCAGAAATAGTCTAGTGGCCCTCTCTGCTCCCTGCTGAA	1122
341	AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro	360
1123	GATGACACCTGTAATCCAGGTGAAGAGGAGGCCAGACCTATGCACCCCTAAGCCG	1182
361	LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProLysAlaMetLys	380
1183	CTGAGTCTCTGACCCAGGACCTCCAGTCTTCCGAAAGTTCTCTTCCAAAACAGCGAAG	1242
381	LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet	400
1243	AAGTTCAGGCGCGCAAGAGAGCTGCGTGGAGTGTCAAGACCGGTGTACCCCATG	1302
401	GluArgLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCys	420
1303	GNAACGGCTCTGGCCCAACAGCAGAGTGTTCACATCAGCTGTTCCGATGCTCTACTGC	1362
421	AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro	440
1363	AAACAAACAGCTCAGTCTAGGAACATATGCATCCTTACATGGACGAATCTACTGTAACT	1422
441	HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro	460
1423	CACCTTCAATCAACTCTTCAAGTCCAAAGGCAACTATGACGAGGGCTTTGGGCACAGCAG	1482
461	HisLysAspLeuTrpAlaSerLysAsnGluAsnGluIleLeuGluArgProAlaGln	480
1483	CATAGAGTATCTGTGGCGAAGCAAGACGCAATATGAGGAGACTTTGGGAGACACGCCAG	1542
481	LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys	500
1543	CGCGCTAATGACGGGACAGCCCCCATAGCCCGGTGTAGAAGATGCCCCCATCGCCAG	1602
501	ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysGluAsp	520
1603	GTCGCGTGTCTGGCGCAAGTATGGAAGCCAAAGGCTCTTCTCAGAGGAGGAGGAGGAT	1662
521	LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer	540
1663	AAGCCCGGAGAACCAAGAACTGAGGATCGCTGCTCGCCACGCGAGCTGGCGGT	1722
541	SerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTyrProProGluAsp	560
1723	TCCGGAATGCTCCGAGAGAGGATCAAACTTCTGAGCTCTCACTGCTCTGCTCTCT	1782

Db 715 AATCTGGAGCTGCCAGCGCTCTCAGAAACCTCCATAAAGAGCCGATGCGCCAGTACCAG 774
Qy 276 AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly 295
Db 775 GCTGCAGTGTCAAGCAGAGCAGCCAGCAGCTACACAAATGAGCTGAAACACAGT--- 831
Qy 296 GlyGluLeuLysIleHisLysMetGluGlnLysGluAsnValProGlyProGluVal 315
Db 832 ---GAAAGCAAAACTCATAAATGGAGCAGAGGAGAAATGCTCCAGGTCCCGAGGCC 888
Qy 316 CysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArg 335
Db 889 TGCACGCTCCATCAGAGAGGAGCAAGGTTCGCACTGAGAAATAGTCTAGTGGCCCTC 948
Qy 336 SerThrProAlaGluAspAspSerArgAspSerGlnValLysSerGluValGlnGlnPro 355
Db 949 TCTGTCCCTGCTGAAGATGACACCTGTAACCTCCAGGTGAAGAGCGAGGCCAGCAGCT 1008
Qy 356 ValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerPro 375
Db 1009 ATGCACCCCTAAGCCCTCAGTCTGACCCAGGACCTCCAGTCTCCGGAAAGTTCTCCT 1068
Qy 376 ProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLys 395
Db 1069 TCCAAACAGCAGAAAGTTTCAGCGCGCGCAAGAGAGAGCTGCTGGAGTGCAGAA 1128
Qy 396 ThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPhe 415
Db 1129 ACGGTGTACCCCATGGAACGCTCTCTGCGCAACAGCAGAGGTGTTTCACATCAGCTGTTT 1188
Qy 416 ArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArg 435
Db 1189 CGATGCTCTCTGCAACAACAGCTAGCTAGTACCAATATGCATCTCTTACATGAGCTTTG 1248
Qy 436 IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGly 455
Db 1249 ATCTACTGTAGCTTCACTTCAATCACTCTTCAAGTCCAAAGGCACTATGACAGCGGC 1308
Qy 456 PheGlyHisArgProHisLysAspLeuTyrAlaSerLysAsnGluAsnGluLeuLeu 475
Db 1309 TTGGGCACAGCAGCATAGATCTGTGGCAGCAGCAGCAGCATGAGGAGACTTTG 1368
Qy 476 GluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAsp 495
Db 1369 GGGAGACCCAGCCGCTTCAATGAGGAGAGCTCCCATACCCCGGTGTAGAGAT 1428
Qy 496 AlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGln 515
Db 1429 GCCCCCATCGCAAGCGCGCTGCTGCGCGCAAGTATGGAAGCAAGGCTCTTCTCAG 1488
Qy 516 GlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTyrProProPro 535
Db 1489 AGGAGAGGAGGATAGCTTGGGAGCAGCAGAGAGCTGAGATCGCTGCTGCTCCGCCA 1548
Qy 536 ThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLys 555
Db 1549 CTTGAGCTGGCAGTTCGGGAAGTGCCTGAGGAGGAGGATCAAGATCGAAGCCCAAG 1608
Qy 556 TrpProProGluAspGluLysSerLysProGluValProGluAspValAspLeuAspLeu 575
Db 1609 TGGCTCCGGAGGATGAGCTGCAAGACGAGGCGCCGAGGATGATGATCTGGACCTG 1668
Qy 576 LysLysLeuArgArgSerSerLeuLysGluArgSerArgProPheThrValAlaAla 595
Db 1669 AAGAGCTGGCGCGCTTCTGCTCCCTGAAGAGAGGAGCGCCCGTTTACTGTAGCAGCT 1728
Qy 596 SerPheGlnSerThrSerValLysSerProLysThrValSerProProIleArgLysGly 615
Db 1729 TCCATTGGAACCTCTTCCATCAGAGCCCAAGACCTCTTCGCCATCTCTCAGGAAGGT 1788
Qy 616 TrpSerMetSerGluGlnSerGluLysSerValGlyGly---ArgValAlaGluArgLys 634

Db 1789 TGGAGCAGTCCGAGCAGAGTGGAGGAGGATGGAGGAGGATGCGAAAGAAA 1848
Qy 635 GlnValGluAsnAlaLysAlaSerLysLysAsnValGlyLysThrThrTrpGln 654
Db 1849 CAAACGGMAATGCCAGACCTCTCGGAGAAAGAGAGATGTGGAAATACACCTGGCAG 1908
Qy 655 AsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMet 674
Db 1909 -----GCCGAGGAAGTGGCAGGAGCAAGGACCGTAGTCTTTTGGAGCTG 1953
Qy 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeu 694
Db 1954 GAGAGTGAGAAATTTTATGGAAATGGAGCAACATAGCTGAAGATGACCAACCATGTCCAC 2013
Qy 695 LysGlnGlnSerProGlnGluProLysSerLeuAsnTyrSerSerPheValAspAsnThr 714
Db 2014 GCACAGCAGTCTCCGTAGAACCCGAGGCTCCAGGTGGCTTGTGTAGACACACG 2073
Qy 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly 734
Db 2074 GCTGCTAAAAGATTCTACTACCCAGAAATCAAAATCCCAAGATGTCGGGTCTCGGAGGGA 2133
Qy 735 GluValValLysGluLeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGlu 754
Db 2134 GAAGTGGTCCGAGAGCTGTCTGTGGAGCAACAGATAAAGAGAAACCGGTACTATGATGAG 2193
Qy 755 AspGluAspGluGlu 759
Db 2194 GACGAGGATGAAGAA 2208

RESULT 11
AF218025 1754 bp mRNA linear PRI 01-OCT-2000
LOCUS Homo sapiens clone PF624 unknown mRNA.
DEFINITION AF218025
ACCESSION AF218025.1 GI:10441979
VERSION FLI CDNA.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1754)
AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,
Qin,W.X., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,
Yu,J. and Han,L.H.
Novel Human cDNA clones with function of inhibiting cancer cell
growth
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1754)
AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,
Yu,J. and Han,L.H.
Direct Submission
Submitted (21-DEC-1999) National Laboratory For Oncogenes & Related
Genes, Shanghai Cancer Institute, 25 Ln 2200, Xie-Tu Road, Shanghai
200032, People's Republic of China
FEATURES
Location/Qualifiers
1..1754
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PF624"
444..1349
/codon_start=1
/product="unknown"
/protein_id="AAG17267.1"
/db_xref="GI:10441980"
/translation="MKFQAPARETCVCQTVTYXMERLLANQQVFHISLCRKYCXN
KLSLGTASLHGRYCKPHFNLFKSGNYDEGFGRPHKDLKASKNENILRRPAQ
LANARETSPGVEDAPIAKGVLAASMEAKSQEKEKPAETKKLRIAPPTTEL
GSSSKALBEIGKMSKPWPEDEISKPEVEDVLDLKLRRSSSLKRSRPFPTVAS
FQSTSKPTVSPPIRKGSMSQSEESVGRVAERKQVENAKSKNGNVGKTTWQ
NKESGEGREVRKVIWVRMRIL"

BASE COUNT 596 a 377 c 426 g 361 t 4 others
ORIGIN

Alignment Scores:

Pred. No.: 3.4e-138 Length: 1754
Score: 2679.00 Matches: 523
Percent Similarity: 98.87% Conservative: 0
Best Local Similarity: 98.87% Mismatches: 5
Query Match: 68.22% Indels: 2
DB: 9 Gaps: 0

US-09-890-549-4 (1-759) x AF218025 (1-1754)

QY 232 AspAspLeuGluLeuGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLysAsn 251
DB 2 GATGACCTGGAATAGCCCGCAGGTGAGTTGTCATCTTACATTTGACCTGGAGAAAAAT 61
QY 252 GluSerArgArgAsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMet 271
DB 62 GAGAGTAGAGCAATCTGGAACTTCCAGCCTCTCAGAACTCTATAAGGATCGAATG 121
QY 272 AlaLysThrGlnAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeu 291
DB 122 GCCAAGTACCAGGCAGCTGTGTCCAAACCAAGCAGCTCAACCAACTATACAAATGAGCTG 181
QY 292 LysAlaSerGlyGluLysLysLysLysLysMetGluGlnLysGluAsnValProPro 311
DB 182 AAGCCAGGTGGTGGCAATCAAAATTCATAAATGGCAAGGAGAGATGTGCCCCCA 241
QY 312 GlyProGluValCysIleThrHisGlnGluGlyGluLysLysLysLysLysLysSer 331
DB 242 GGTCTCTGAGCTCTGCATCACCCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 301
QY 332 LeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnValLysSerGlu 351
DB 302 CTGGCAGTCCGTCTCCACCCTGCGAAGATGACTCCCGTCTCCAGCTTAAAGAGTGAG 361
QY 352 ValGlnGlnProValHisProLysProLysProLysProLysProLysProLysSer 371
DB 362 GTTCAACAGCCTGTCCATCCCAAGCCCACTAAGTCCAGATCTCCAGAGCCTCCAGTCTTTC 421
QY 371 rGluSerProProLysAlaMetLysLysLysPheGlnAlaProAlaArgGluThrCysVa 391
DB 422 TGNAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGT 481
QY 391 lGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHi 411
DB 482 GGAATGTCAGAGACAGTCTATCAATGAGCGTCTCTTGGCCACCCAGCAGGTGTTTCA 541
QY 411 sIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSe 431
DB 542 CATCAGCTGTCRCGTGTCRCATATGCAKCAACAACTCAGTCTAGGAACATATGCATC 601
QY 431 rLeuHisGlyArgGlyLeuTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAs 451
DB 602 TTTACATGGAAGAATCTATTGAAGCCTCACTTCAATCAACTCTTAAATCTAAGGGCAA 661
QY 451 nTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLysAsnGluAs 471
DB 662 CTATGATGAAGGCTTTGGCCACAGCCACCAAGGATCTATGGGCAAGCAAAATGAAGA 721
QY 471 nGluGluLeuLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerPr 491
DB 722 CGAAGAGATTTTGGAGAGACCCAGCCAGCTTCAATGCAAGGGAGAGCCCTCACAGGCC 781
QY 491 oglyValGluAspAlaProIleAlaLysValGlyValLeuAlaLaSerMetGluAlaLy 511
DB 782 AGGGGTAGAAGATGCCCTATTGCTAAGGTGGCTCTCTGGCTGCAAGATGATGAAGCCAA 841
QY 511 sAlaSerSerGlnGlnGluLysGluAspLysProAlaGlnThrLysLysLeuArgIleAl 531
DB 842 GGCCTCTCTCAGCAGGAGGAGGAGCAAGCCAGCTGAACCAAGAGAGCTGAGGATCGC 901

QY 531 aTTPProProThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyIleLysMe 551
DB 902 CTGGCCACCCCTGAACTGAACTTGGAAAGTTCAGGAAGTCTGGAGGAAGGATCAAAAT 961
QY 551 tSerLysProLysTyrProProGluAspGluLysSerLysProGluValProGluAspVa 571
DB 962 GTCAAGAGCCCAATGCGCTCTCTGAAGAGCAAAATCAGCAAGCCCGCAAGTCTCTGAGATGT 1021
QY 571 lAspLeuAspLeuLysLysLeuArgSerSerSerSerLeuLysGluArgSerArgProPh 591
DB 1022 CGATCTAGATCTGAAGAAGCTAAGACGATCTTCTCAAGGAAGAAAGCCGCCAT 1081
QY 591 eThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPr 611
DB 1082 CACTCTAGCAGCTTCAATTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTGTCTCCACC 1141
QY 611 oileArgLysGlyTyrSerMetSerGluGlnSerGluSerValGlyArgValAl 631
DB 1142 TATCAGGAAGAGCTGAGCATGTGAGCAGCAGTGAAGAGTCTGTGGGTGGAAGAGTTGC 1201
QY 631 aGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysTh 651
DB 1202 AGAAGGAAACAAGTGGAAATGCCAAGCTTCTAAGGAAGATGGAAATGTGGGAAAAAC 1261
QY 651 rThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSe 671
DB 1262 AACCTGGCAAAACAAAGAAATCTAAAGGAGAGCA -GGGAAGAGAGTAAGGAAGTCTAG 1320
QY 671 rLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAs 691
DB 1321 TTTGAGATGGAGAAATGAGAAATCTTTGTAGAAAATGGTCAGACTCCGATGAAGATGATA 1380
QY 691 nSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheVa 711
DB 1381 CAGCTTCTCCAAACAACAATCTCCACAAGAACCAAGTCTCTGAATTTGTCGAGTTTGT 1440
QY 711 lAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLe 731
DB 1441 AGACACACCTTTTGTCTGAGAAATTCATCTACTCAGATCAGAAATCCCAAGGATGGAACT 1500
QY 731 uTPGluGlyGluValValLysGluLeuSerValGluGluGlnLysLysArgAsnArgTy 751
DB 1501 CTGGAGGAGGAGAGTGGTCAAAGAGCTCTCTGTGGAAGAACAGATAAGAGAAATCGTA 1560
QY 751 rTyrAspGluAspGluAspGluGlu 759
DB 1561 TTATGATGAGGATGAGGATGAAGAG 1585
RESULT 12
BD160117
LOCUS BD160117 2207 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD160117
VERSION BD160117.1 GI:27865875
KEYWORDS JP 2002191363-A/14960.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2207)
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primer for synthesizing full-length cDNA and use thereof
TITLE Patent: JP 2002191363-A 14960 09-JUL-2002;
JOURNAL HELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/14960
PD 09-JUL-2002
PF 28-JUL-2000 JP 200280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
KEIICHI NAGAI, TETSUO IOTSUKI

PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
10,
PC C12P21/02,C12Q1/69/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS (580)..(2050).
Location/Qualifiers
1..2207
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 692 a 438 c 512 g 565 t
ORIGIN
Alignment Scores:
Pred. No.: 2,37e-126 Length: 2207
Score: 2465.00 Matches: 471
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 62.77% Indels: 0
DB: 6 Gaps: 0
US-09-890-549-4 (1-759) x BD160117 (1-2207)
QY 289 AsnGluLeuLysAlaSerGlyGluLeuLysHisLysMetGluGlnLysGluAsn 308
Db 638 AATGAGCTGAAGGCGAGTGGTGGCAATCAAAATTCATAAAATGGAGCAAGAGAAAT 697
QY 309 ValProGlyProGluValCysIleThrHisGlnGluGlyLysIleSerAlaAsn 328
Db 698 GTGCCCCAGGCTCTGAGGTCTGATCACCCTCAGGAGGAGGAAAGATTTCTGCAAAAT 757
QY 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnVal 348
Db 758 GAGATAGCTGGAGTCCGTTCCACCCCTGCCAGAGATGACTCCCGTGACTCCAGGTT 817
QY 349 LysSerGluValGlnGlnProValHisProLysProLeuSerProAspSerArgAlaSer 368
Db 818 AAGAGTGAGGTTCAACAGCCTGTCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 877
QY 369 SerLeuSerGluSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388
Db 878 AGTCTTCTGAAGATTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 937
QY 389 ThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGln 408
Db 938 ACCTGCGTGGNATGTCAGAGACAGTCTATCCATGGAGCGTCTCTCTGGCCAAACCCAGCAG 997
QY 409 ValPheHisLysSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThr 428
Db 998 GTGTTTCAATCAGCTGCTTCGTTGCTCTCTATTCGCAACCAACTCAGTCTAGGAACA 1057
QY 429 TyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448
Db 1058 TATGCATCTTTACATGGAAGAAATCTATTAAGGCTCATTCAATCAACTCTTTAAATCT 1117
QY 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLys 468
Db 1118 AAGGCAACTATGATGAAGCTTTGGGCAAGACCCACAGAGGATCTATGGGCAAGCAAA 1177
QY 469 AsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 1178 AATGAAACCAAGAGATTTTGGAGAGACCCAGCCAGCTTCAATGCAAGGGAGACCCCT 1237
QY 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaSerMet 508
Db 1238 CACAGCCCGGGGTAGAAGATGCCCCCTATTGCTAAGGTGGGTCTCTGGCTGCAAGTATG 1297
QY 509 GluAlaLysAlaSerSerGlnGlnGluLysGluAspLysProAlaGluThrLysLysLeu 528
Db 1298 GAAGCCAAAGGCTCTCTCAGCAGGAGAGGAGAGCAAGCCAGCTGAAACCAAGAGCTG 1357

QY 529 ArgIleAlaTrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548
Db 1358 AGGATCGCTGGCCACCCCTTGAAGTTCAGGAAGTCCCTTGGAGGAAGG 1417
QY 549 IleLysMetSerLysProLysTrpProProGluAspGluLysSerLysProGluValPro 568
Db 1418 ATCAAAATGTCAAGCCCAATGGCTCTCTGAAGACGAATCAGCAAGCCCAAGTTCT 1477
QY 569 GluAspValAspLeuAspLeuLysLysLeuArgSerSerSerLeuLysGluArgSer 588
Db 1478 GAGGATCTCGATCTAGATCTGAAGAGCTTAAGACCATCTTCTTCTACTGAAGAAAGAAC 1537
QY 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 1538 CGCCCATTCACCTAGCAGCTTCATTTCAAGCACCTCTGTCAAGAGCCCAAACTGTG 1597
QY 609 SerProProIleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGly 628
Db 1598 TCCCCACCTATCAGGAAGGCTGGAGCATGTCAAGACAGAGTGAAGAGTCTCTGGGTGGA 1657
QY 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648
Db 1658 AGAGTTGCAGAAAGGAAACAAGTGGAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTG 1717
QY 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 1718 GGAAACCAACCTTGGCAAAACAAGATCTAAAGAGAGACAGGAGAGAAAGTAAGGAA 1777
QY 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 1778 GGTCTAGTTTGGAGATGGAGATGAGAATCTTTAGAAAATGGTGCAGACTCCGATGAA 1837
QY 689 AspAspAsnSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSer 708
Db 1838 GATGATAACAGCTTCTCTCAACCAACAATCTCCACAAAGAACCCCAAGTCTCTGAATTGGTGC 1897
QY 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAsp 728
Db 1898 AGTTTGTAGACACACCTTTCTCTGAAGATTCATCTACTCAGAAATCAGAAATCCCAGGAT 1957
QY 729 ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGluGlnLysArg 748
Db 1958 GTGGAATCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGACAGATAAGAGA 2017
QY 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 2018 AATCGTATTATGATGAGGATGAGGATGAAGAG 2050
RESULT 13
LOCUS AK023649 2207 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ13587 fis, clone PLACE1009246, weakly similar
to POLLEN SPECIFIC PROTEIN SF3.
ACCESSION AK023649
VERSION AK023649.1 GI:10435634
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hoshino,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Iehida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahara,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
NEO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2207)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission

JOURNAL Submitted (23-AUG-2000) Takao Iwagai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@ri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES Location/Qualifiers

source 1.2207

organism="Homo sapiens"

mol_type="mRNA"

db_xref="taxon:9606"

clone="PLACE1009246"

tissue_type="placenta"

clone_lib="PLACE1"

notes="cloning vector: pME18SFL3"

680..2053

note="unnamed protein product"

codon_start=1

protein_id="BA014625.1"

db_xref="GI:10435635"

/translation="MEQKENVPPGVEVCIHQEGEKISANENSLAVRSTPADDSDRDS QVKEVQOPVHPKPLSPDRASSLSSESPPKAMKFOAPARETCVCEQKTVYMERLL ANQOVPHISCFRCYSYNNKLSLCTYASLHCRIVYKXPHQLFKSKNYDEGRGHRPHK DLWAKNNEEILERAQLANARETHPSGVEDAPLAKGVLAASWEAKSSQKBEKED KPAETKKLRITAMPPTIELSSGSALEEGINMSKPKWPPDEISKEPVEDVDLKLK RRSLSKERSPTTAAQSTSVKSPKTVPPPIRKGMNSQSEESVGRVAERKQV ENAKASKQNGNVGKTQWKEKSGTGKSKGHESELENNENLVENGADSDDDNSFL KOQSPQPKSLNKSFFVDNFTAFEEFTTQNKSDVLEWGEVVKLSVEQIKRNYR DEDEDEE"

BASE COUNT 692 a 438 c 512 g 565 t

ORIGIN

Alignment Scores:

Pred. No.: 2,37e-126 Length: 2207

Score: 2455.00 Matches: 471

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 62.77% Indels: 0

DB: 9 Gaps: 0

US-09-890-549-4 (1-759) x AK023649 (1-2207)

289 AsnGluLeuLysAlaSerGlyGlyGluIleLysHisLysMetGluGlnLysGluAsn 308

638 AATGAGCTGAACCCAGTGTGGCGAAATCAAAATTCATAAATGAGCAAGAGGAGAT 697

309 ValProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsn 328

698 GTGCCCCCAGGTCTGAGGTCTGCATCACCCATCAGGAGGGGAAAGATTCTGCAAT 757

329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAlaSer 348

758 GAGAAATAGCTGGCATCGTTCACCCCTGCGAAGATGACTCCGCTACCTCCAGGTT 817

349 LysSerGluValGlnGlnProValHisProLysProLysProLysSerArgAlaSer 368

818 AAGAGTGAGGTTCAACAGAGCTGTCCATCCAGCCACTAAGTCCAGATTCCAGAGCTCC 877

369 SerLeuSerGluSerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388

878 AGTCTTCTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTCCAGAGAG 937

389 ThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsnGln 408

938 ACTCGGTGGAATGTCAGAAGACAGTCTATCCAAATGGAGCGTCTCTGSCCAACCCAGCAG 997

409 ValPheHisIleSerCysPheArgCysSerTyCysAsnLysLeuSerLeuGlyThr 428

998 GTGTTTCACATCAGCTGCTTTCGTTGCTCTTATTCACCAACAACTCAGTCTAGGAACA 1057

429 TyrAlaSerLeuHisGlyArgIleTyrcysLysProHisPheAsnGlnLeuPheLysSer 448

1058 TATGCACTTTACATGGAAGAATCTATTGTAAAGCTTCACTTCAATCAACTCTTTAAATCT 1117

449 LysGlyAsnTyrcysGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLys 468

1118 AAGGCACTATGATGAAGCTTTGGGCACACACACCAAGGTTCTATGGGCAAGCAAA 1177

469 AsnGluAsnGluGluLeuLeuArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488

1178 AATGAAACCAAGAGATTTTGGAGAGACCCAGCCAGCTTGCATTAAGGAGAGAGCCCT 1237

489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaSerMet 508

1238 CACAGCCAGGCTTGAAGATGCCCTTATGTAAAGTGGGTGCTCTGGCTCAAGATG 1297

509 GluAlaLysAlaSerSerGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528

1298 GAAGCAAGGCTCTCTCAGCGAGAGGAAGCAAGCCAGCTGAAACCAAGAGAGCTG 1357

529 ArgIleAlaTyrProProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548

1358 AGGATCGCTGGCCACCCCTTGAAGTTCAGGAGTTCAGGAGTTCCTTGGAGAGAGG 1417

549 IleLysMetSerLysProLysTrpProGluAspGluIleSerLysProGluValPro 568

1418 ATCAAAATGTCAAGCCCAATGGCTCTCTGAAGACGAATCAGCAAGCCGAGTCTCT 1477

569 GluAspValAspLeuAspLeuLysLysLeuArgArgSerSerLeuLysGluArgSer 588

1478 GAGGATGCGATCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGAGAGAGC 1537

589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608

1538 CGCCCATTTCACTGTAGCAGCTTCAATTCAGACCTCTGTCAAGAGCCCAAACTGTG 1597

609 SerProIleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGly 628

1598 TCCCCACCTATCAGAAAGGCTGGAGCATGTGAGAGAGTGAAGAGTCTGTGGGTGA 1657

629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnVal 648

1658 AGAGTTGAG 1717

649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668

1718 GGAAG 1777

669 GlyHisSerLeuGluMetGluAsnGluValGluAsnGlyAlaAspSerAspGlu 688

1778 GGTATAGTTTGGAGATGAGAGATGAGATCTTTAGAGAGAGAGAGAGAGAGAGAGAG 1837

689 AspAsnSerPheLeuLysGlnSerProGlnSerProGlnSerProLysSerLeuAsnTrpSer 708

1838 GATGATAACAGCTTCTTCAACCAACATCTCCACAGAGAGAGAGAGAGAGAGAGAGAG 1897

709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAsp 728

1898 AGTTTGTAGACACACCTTTGCTGAAGAGATTTACTACTCAGAAATCAGAAATCCAGAT 1957

729 ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArg 748

1958 GTGAAGCTCTGGAG 2017

749 AsnArgTyrcysAspGluAspGluGluGlu 759

2018 AATCGGTATTATGATGAGGATGAGGATGAGAG 2050

RESULT 14

AK096172

LOCUS

DEFINITION Homo sapiens cDNA FLJ38853 fis, clone MESAN2010321, highly similar

2793 bp mRNA linear PRI 15-JUL-2002

to Homo sapiens epithelial protein lost in neoplasm beta (EPLIN)
mRNA.
ACCESSION AK096172
VERSION AK096172.1 GI:21755594
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yanamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 2793)
REFERENCE Isogai,T. and Yamamoto,J.
AUTHORS Direct Submission
TITLE Submitted (04-JUL-2002) Takao Isogai, FUJ Project (HRI Team); 2-6-7
JOURNAL Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5' - & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1. 2793
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MESAN2010321"
/cell_type="normal mesangial cells (NHMC56046-2)"
/clone_lib="MESAN2"
/note="Cloning vector: pME18SFL3-primary culture, normal
mesangial cells"
BASE COUNT 826 a 577 c 637 g 753 t
ORIGIN
Alignment Scores:
Pred. No.: 5,89e-126 Length: 2793
Score: 2460.00 Matches: 470
Percent Similarity: 99.79% Conservative: 0
Best Local Similarity: 99.79% Mismatches: 1
Query Match: 62.64% Indels: 0
DB: 9 Gaps: 0
US-09-890-549-4 (1-759) x AK096172 (1-2793)
Qy 289 AsnGluLeuLysAlaSerGlyGlyGluLeuLysIleHisLysMetGluGlnLysGluAsn 308
Db 1230 AATGAGCTGAAAGCCAGGTGGCGAATCAATTCATAAAATGGAGCAAGAGGAAT 1289
Qy 309 ValProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsn 328
Db 1290 GTGCCCCAGGTCTGAGGTCTGATCCATCCATCAGGAAGGGGAAAGATTTCTGCAAT 1349
Qy 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnVal 348
Db 1350 GAGAAATAGCTGGGAGCTCGCTCCACCCCTCCCGAAGATGACTCCCGTACTCCCAAGTT 1409
Qy 349 LysSerGluValGlnGlnProValHisProLysProLeuSerProAspSerArgAlaSer 368
Db 1410 AAGAGTGAGGTTCACAGCTGTTCATCCCAAGCCATAGTCCAGATTCCAGAGCCCTCC 1469
Qy 369 SerLeuSerGluSerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388

Db 1470 AGTCTTTCTGAAGTTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTCCAGAGAG 1529
Qy 389 ThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsnGlnGln 408
Db 1530 ACCTGGTGGATGTGTCAGAGACAGTCTATCAATGAGCGTCTCTTGCCCAACAGCAG 1589
Qy 409 ValPheHisIleSerCysPheArgCysSerTyCysAsnAsnLysLeuSerLeuGlyThr 428
Db 1590 GTGTTTTCATCAGTCTGCTTCCGTTCTCTTATGTCATTCGCAACAACTCAGTCTAG 1649
Qy 429 TyrAlaSerLeuHisGlyArgIleTyCysValProHisPheAsnGlnLeuPheLysSer 448
Db 1650 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAA 1709
Qy 449 LysGlyAsnTyAspGluGlyPheGlyHisArgProHisLysAspLeuThrAlaSerLys 468
Db 1710 AAGGGCAACTATGATGAGGCTTTGGGCACACACACCAAGATCTATGGGCAAGCAA 1769
Qy 469 AsnGluAsnGluGluLeuLeuArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 1770 AATGAAACGAAGAGATTTTGGAGAGACAGCCAGCTTGCATAATGCAAGGAGAGCCCT 1829
Qy 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaLaserMet 508
Db 1830 CACAGCCAGGGGTAGAAGATGCCCTTATGCTAAGGTGGGTCTCTGCTGCTCAAGAT 1889
Qy 509 GluAlaLysAlaSerSerGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528
Db 1890 GAAGCCCAAGCCTCTCTCAGCAGGAGAAAGAGCAAGCCAGCTGAAACCAAGAGCTG 1949
Qy 529 ArgIleAlaTrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548
Db 1950 AGGATCGCTGGCCACCCCTCACTGAAGTTCAGGAGTGCCTTGGAGGAGGG 2009
Qy 549 IleLysMetSerLysProLysTrpProGluAspGluLysSerLysProGluValPro 568
Db 2010 ATCAAAATGTCAAGCCCAATATGGCTCTCTGAAGACGAATCAGCAAGGCCCAAGTTCT 2069
Qy 569 GluAspValAspLeuAspLeuLysLysLeuArgSerSerSerLysLysGluArgSer 588
Db 2070 GAGGATGTCTGATCTAGTCTGAAGAGCTAAGACCACTTCTTCACTGAAGAAAGAAC 2129
Qy 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 2130 CGCCCATTCAGTGTAGCAGCTTCATTTCAAGCACCTCTCTCAAGAGCCCAAAACTGTG 2189
Qy 609 SerProProIleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGly 628
Db 2190 TCCCACTATCAGGAAAGGCTGGAGCATGTCAGAGCAGAGTGAAGAGTCTCTGGGTGA 2249
Qy 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648
Db 2250 AGAGTTGCAGAAAGGAAACAGTGGAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTG 2309
Qy 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 2310 GGAATAAACCTCGCAAAACAAAGAAATCTAAAGGAGAGACAGGGAAGAGTGAAGAA 2369
Qy 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 2370 GGTGATAGTTGAGATGGAGATCAGAAATCTTTAGAAAATGGTGCAGACTCCGATGAA 2429
Qy 689 AspAspAsnSerPheLysGlnGlnSerProGlnGluProLysSerLysLeuAsnTrpSer 708
Db 2430 GATGATAACAGCTTCTCAACCAACAACTCTCCCAAGAACCCCAAGTCTCTGAATTTGGT 2489
Qy 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAsp 728
Db 2490 AGTTTGTAGACAAACACCTTTCTGTAAGAAATTCAGTCTCAGAAATCAGAAATCCCAG 2549
Qy 729 ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGlnLysArg 748

Db 2550 GTGGAACCTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAGAGA 2609

Qy 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759

Db 2610 AATCGGTATTATGATGAGGATGAGATGAAGAG 2642

RESULT 15

AX017487

LOCUS AX017487 2783 bp DNA linear PAT 07-SEP-2000

DEFINITION Sequence 31 from Patent WO9947655.

ACCESSION AX017487

VERSION AX017487.1 GI:10042284

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and Pilarsky,C.

TITLE Human nucleic acid sequences from normal breast tissue

JOURNAL Patent: WO 9947655-A 31 23-SEP-1999;

SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN (DE); PILARSKY CHRISTIAN (DE)

FEATURES

Location/Qualifiers

1..2783

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

BASE COUNT 901 a 526 c 609 g 747 t

ORIGIN

Alignment Scores:

Pred. No.: 7,55e-126 Length: 2783

Score: 2458.00 Matches: 470

Percent Similarity: 99.79% Conservative: 0

Best Local Similarity: 99.79% Mismatches: 1

Query Match: 62.59% Indels: 0

DB: 6 Gaps: 0

US-09-890-549-4 (1-759) x AX017487 (1-2783)

Qy 289 AsnGluLeuLysAlaSerGlyGlyGluLeuLysLeuLysMetGluGlnLysGluAsn 308

Db 40 AATGAGCTGAAGCCAGTGGTGGCAATCAAAATTCATAAATGGAGCAAGAGAGAT 99

Qy 309 ValProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsn 328

Db 100 GTCCCCCAGCTCCTGAGGTCTGCATCCCATCCCATCAGGAAGGGGAAAGATTTCTGCAAT 159

Qy 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnVal 348

Db 160 GAGAAATAGCTGCGAGTCCGTTCCACCCCTGCCGAGATGACTCCGTCGACTCCAGGTT 219

Qy 349 LysSerGluValGlnGlnProValHisProLysProLysProAspSerArgAlaSer 368

Db 220 AAGAGTGAGGTTCAACAGCGCTGTCCATCCCAAGCCACTAAGTCCAGATCCAGAGCTCC 279

Qy 369 SerLeuSerGluSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388

Db 280 AGTCTTTCTGAAAGTTCTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTCAAGAGAG 339

Qy 389 ThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGln 408

Db 340 ACCTGCGTGGATGTGAGAAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCCAACCAGAG 399

Qy 409 ValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThr 428

Db 400 GTGTTTCACATCAGCTGCTTCGGTTGCTCTATTTGCAACAACAACACTCAGTCTAGAACAC 459

Qy 429 TyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448

Db 460 TATGCATCTTTACATGGAAGAATCTATTTCATAGCCCTCACTTCAATCAACTCTTTAAATCT 519

Qy 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLys 468

Db 520 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACACAGAGATCTATGGGCAAGCAA 579

Qy 469 AsnGluAsnGluGluLeuLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488

Db 580 AATGAAACGAAGAGATTTTGAGAGAGACAGCCAGCTTGCAAATGCAAGGGAGACCCCT 639

Qy 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMet 508

Db 640 CACAGCCAGGGGTAGAAGATGCCCTATTGCTAAGGGGGGTCTCTGGCTCAAGTAGT 699

Qy 509 GluAlaLysAlaSerSerGlnGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528

Db 700 GAAGCAAGGCTCTCTCAGCAGAGAGGAAGCAAGCCAGCTGAAACCAAGAGCTG 759

Qy 529 ArgIleAlaTrpProProThrProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548

Db 760 AGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTTCAGGAAGTGCCTTGGAGGAAG 819

Qy 549 IleLysMetSerLysProLysTrpProGluAspGluIleSerLysProGluValPro 568

Db 820 ATCAAAATGTCAAGCCCAATGGCTCTCTGAAGACGAATCAGCAAGCCCAAGTTCTCT 879

Qy 569 GluAspValAspLeuAspLysLysLysLeuArgArgSerSerSerLeuLysGluArgSer 588

Db 880 GAGGATGTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAGAAG 939

Qy 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608

Db 940 CGCCCATTCACCTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAACTGT 999

Qy 609 SerProProIleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGly 628

Db 1000 TCCCCACCTATCAGGAAAGGCTGGAGCATGTCAAGCAGAGTGAAGAGTCTGTGGTGA 1059

Qy 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648

Db 1060 AGAGTTGCAAGAAAGAAACAAGTGGAAATGCCAAGGCTTCTAAGAAATGGGAATGT 1119

Qy 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668

Db 1120 GGAAAAACAACCTGGCAAAACAAGAAATCTAAGAGAGACAGGAGAGAGAGTAAGGAA 1179

Qy 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688

Db 1180 GGTCAATAGTTGGAGATGGAGATCAGAAATCTTGAGAAATGGTGACAGACTCCGATGA 1239

Qy 689 AspAspAsnSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSer 708

Db 1240 GATGATAACAGCTTCTCTCAACAAACAATCTCCCAAGAACCCCAAGTCTCTGAATTGGT 1299

Qy 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGluAsp 728

Db 1300 AGTTTGTAGACAACACTTTGCTGAAGAAATTCCTACTCAGAAATCAGAAATCCAGAT 1359

Qy 729 ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArg 748

Db 1360 GTGGAATCTCTGGAGGAGAGAGTGGTCAAGAGAGCTCTCTGTGGNAGAACAGATAAGAGA 1419

Qy 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759

Db 1420 AATCGTATTATGATGAGGATGAGGATGAAGAG 1452

Search completed: January 7, 2004, 00:38:49

Job time : 7272 secs

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
------------	-------	-------------	--------	----	----	-------------

1	3927	100.0	2905	22	AAI58169	Human polynucleoti
2	3927	100.0	3664	23	ABX71307	Human transcrip
3	3927	100.0	3705	21	AAAS3826	Sequence encoding
4	3909.5	99.6	3711	21	AAQ77143	Human ORFX ORF2698
5	3872.5	98.6	3650	22	AAFS5697	Human EPLIN (epith
6	3603.5	91.8	3543	22	AAFS5696	Human EPLIN (epith
7	3285.5	83.7	3465	23	AAI58169	DNA encoding novel
8	2679	68.2	1754	24	AB198873	Human cancer supp
9	2537	64.6	1713	22	AAI58169	Human polynucleoti
10	2535	64.6	2749	24	ABX35261	Human cDNA encodin
11	2465	62.8	2207	22	AAH18425	Human cDNA sequenc
12	2455	62.5	2783	22	AAI58169	Human breast tumou
13	2449	62.4	1567	24	AAI58169	CDNA encoding huma
14	1586	40.4	2267	21	AAI58169	Human secreted pro
15	1411	35.9	2158	21	AAI58169	Human secreted pro
16	1254	31.9	732	21	AAI58169	Human pancreatic c
17	881	22.4	641	20	AAI58169	Human prostate tum
18	841.5	21.4	698	22	AAH72864	Human cervical can
19	828	21.1	565	24	ABV98930	Human pancreatic c
20	815	20.8	1014	23	AAI58169	DNA encoding novel
21	615.5	15.7	4269	24	ABQ61002	Human breast tumou
22	606	15.4	413	25	AAI58169	Chinese hamster si
23	596.5	15.2	3825	24	AAI58169	Human lung specifi
24	541	13.8	2547	22	AAI58169	DNA encoding human
25	538	13.7	2336	22	AAI58169	DNA encoding human
26	534.5	13.6	2379	24	AAI58169	Human DNA #13 core
27	534.5	13.6	2379	24	AAI58169	Human 41441 protei
28	506.5	12.9	577	23	ABV49283	Human prostate exp
29	442.5	11.3	710	25	AAI58169	Chinese hamster si
30	423	10.8	532	22	AAI58169	Human cDNA clone (
31	414	10.5	389	21	AAI58169	Human secreted pro
32	392	10.0	1638	24	ABI70363	Human bone remodel
33	390	9.9	1848	23	AAI58169	DNA encoding novel
34	388.5	9.9	1321	24	ABI39708	Human NS cDNA sequ
35	385.5	9.8	2226	23	AAI58169	DNA encoding novel
36	385	9.8	1298	21	AAI58169	Human inflammation
37	384	9.8	408	23	AAI58169	DNA encoding novel
38	383	9.8	1122	24	ABI39709	Human NS cDNA sequ
39	383	9.8	1205	24	ABI39710	Human NS cDNA sequ
40	383	9.8	1284	24	ABI39707	Human NS cDNA sequ
41	379.5	9.7	1159	24	ABI39712	Human NS cDNA sequ
42	379.5	9.7	1242	24	ABI39711	Human NS cDNA sequ
43	374.5	9.5	654	25	AAI58169	Chinese hamster si
44	372	9.5	847	23	AAI58169	DNA encoding novel
45	370.5	9.4	1503	22	AAI58169	Human immune/haema

ALIGNMENTS

RESULT 1
AAI58169
ID AAI58169 standard; cDNA; 2905 BP.
XX
AC AAI58169;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 372.
XX

Human; neotrophic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokine; thrombolytic; drug screening; arthritis; inflammation;
leukaemia; ss.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 6, 2004, 16:28:21 : Search time 497 Seconds
(without alignments)
4122.482 Million cell updates/sec

Title: US-09-890-549-4
Perfect score: 3927
Sequence: 1 MESSFPNRRQTSLSLRVTA.....LSVBEQIKRNPYDEDEDEE 759

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 252756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US0980549/runat.06012004.094752.19815/app.query.fasta.1.903
-DB=N.Geneseq.19Jun03 -QMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=closm62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=prc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0980549 @CGN 1.1.312@runat.06012004.094752.19815 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.19Jun03.*
1: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID1/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
23: /SID1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
24: /SID1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*
25: /SID1/gcgdata/geneseq/geneseqn-emb1/NA2004.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

OS Homo sapiens.
XX WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR P-PSDB, AAM39013.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Claim 1; SEQ ID NO 372; 10078pp; English.
XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX SQ Sequence 2905 BP; 965 A; 617 C; 691 G; 632 T; 0 other;

Alignment Scores:
Pred. No.: 1.25e-272 Length: 2905
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-890-549-4 (1-759) x AAI58169 (1-2905)

QY 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 173 ATGGAAATCATCTCCATTAAATACAGCGCAATGACCTCACTATCATTTGAGGGTAACAGCC 232

QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerTys 40
Db 233 AAGAAGACTTCTCTTCTCAACAAGAAACAAGTCAATCGGCTATTGTGGAATATTCTCCAAG 292

QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 293 TACCAGAAAGCAGCTGAGAGAAACAACATCGGAGAAAGAGAGAAAGTAACCCGAAATCTC 352

QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80

Db 353 TCCAGCACTTTAGAAAGGGGACCTGACTGTGTTAAAGAAAGAGTGGGAGAACCCAGGG 412
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 413 CTGGGAGCAGAGTCTCACACAGACTCTCTACCGAAGCAGCAGCAGTATAGGCACAGA 472
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
Db 473 GCAGACCATCTCTCTGCTGAAGTGACAGCCACGCTGCTTCTGGAGCCAAAGCTACCAA 532
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 533 GAAGAACAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAGGCTCTCTCAGGCT 592
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 593 CGATATCCCCACATCAAGACCGTCAAGATCTTAAAGACCACTCAACAGAAAGTAAAAA 652
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 653 ATGGAAATTTGCTAGGAGATCCAGGATGAAGTAGAAATCAGAAATCAGTGAATAC 712
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db 713 ACAGATGCTTCGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAGATGATG 772
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
Db 773 TTTGAGAAAGGTGAACCAACTCAACTAAGATTCTCCGGGCCCAAGCCGAAGTCAAGT 832
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
Db 833 GGAAGAGAGATCTCGAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCAGGTGAG 892
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db 893 TTGTCATCTTCTACATTTGACTCGAGAGAAATAGAGTAGACGAAATCTGGAATTTCCA 952
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db 953 CGCCTCTCAGAAACCTCTATAAGGATCGAATGGCAAGTACCAGGAGCTGTGTCCAA 1012
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLysLysAlaSerGlyGlyGluLysIle 300
Db 1013 CAAGACAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGTGGCAAAATCAA 1072
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db 1073 CATAAATGGAGCAAAAGAGAAATGTGCCCCAGGTCTCTGAGGTCTGCATCACCATCAG 1132
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 1133 GAAGGGGAAAAGATTTCTGCAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCTGCCGAA 1192
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
Db 1193 GATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGACCTGTCATCCCAAGCCA 1252
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1253 CTAAGTCAGATTCAGAGCCTCCAGTCTTCTGAAAGTTCTCTCCCAAGCAATGAAG 1312
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
Db 1313 AAGTTTCAGGCACCTCGAGAGACCTCGGTGAATGTGAGAGAGAGAGTATATCCAAATG 1372
QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
Db 1373 GAGCGTCTCTGGCCAAACAGCAGTGTTCACATCAGTGTCTCCGTTCCGTTCTCTATTGC 1432
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440

1433 AACAAACAACACTCAGCTAGGAACATATGCATCTTTACATGGAGAAATCTATTGTGAAGCCT 1492
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1493 CACTTCAATCAACTCTTTAAATCTTAAGGCAACTATGATGAAGGCTTTGGGCACAGACCA 1552
QY 461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluLysLeuGluArgProAlaGln 480
Db 1553 CACAAGGATCTATGGCAGCAAGCAAAATGAAACGAGAGATTTTGGAGAGACCCAGCCAG 1612
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1613 CTTGCAATGCAAGGAGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTTATTGCTTAAG 1672
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysGluAsp 520
Db 1673 GTGGGTCTCTGGCTGCAAGTATGAGCAAGGCTCTCTCAGCAGAGAGAGAGAC 1732
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
Db 1733 AAGCAGCTGAAACCAAGAACTGAGGATCGCTGGCCACCCCTGAACTTGAAGT 1792
QY 541 SerGlySerAlaLeuGluGluGlyLysMetSerLysProLysTrpProProGluAsp 560
Db 1793 TCAGGAAGTCTCTGGAGAGAGGATCAAAATGTCAAAGCCCAATGGCTCTCTGAAGAC 1852
QY 561 GluLeSerLysProGluValProGluAspValAspLeuLysLysLysLeuArgArg 580
Db 1853 GAAATCAGCAAGCCGAGTCTCTGAGGATGTCGATCTAGATCTGAAGAGCTTAAGACGA 1912
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
Db 1913 TCTTCTTCACTGAAGGAAAGAGCGCCCTTCACTGTAGCAGCTTCAATTCAAAGCACC 1972
QY 601 SerValLysSerProLysThrValSerProIleArgLysGlyTrpSerMetSerGlu 620
Db 1973 TCTGTCAAGAGCCCAAAACCTGTGTCCCACTATCAGGAAGGCTGGAGCATCTCAGAG 2032
QY 621 GlnSerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
Db 2033 CAGAGTGAAGAGTCTGTGGTGGAGAGGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAG 2092
QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrTrpGlnAsnLysGluSerLysGly 660
Db 2093 GCTTCTAAGAGAATGGGAATGTGGGAAAAACAACCTGGCAAAAACAAAGAACTTAAGGA 2152
QY 661 GluThrGlyLysArgSerLysGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 2153 GAGACAGGGAAGAGAGTAAGCAAGGTATAGTTGGAGATGGAGAAATGAGAAATCTTGT 2212
QY 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGln 700
Db 2213 GAAATGTGTGAGACTCCGATGAAGATGATTAACAGCTTCTCTCAAAACAACTCTCCACA 2272
QY 701 GluProLysSerLeuAsnTrpSerPheValAspAsnThrPheAlaGluGluPheThr 720
Db 2273 GAACCAAGTCTCTGATTTGGTCAGTTTGTAGACAAACCTTTGCTGAAGAAATTCAC 2332
QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu 740
Db 2333 ACTCAGAATCAGAAATCCAGGATGTGGAATCTGGCAGGAGAGAGTGTGCAAGAGCTC 2392
QY 741 SerValGluGluGlnLysArgAsnArgTyrTyrAspGluAspGluGlu 759
Db 2393 TCTGTGGAAGAACAGATAAGAGAAATCGGTATTATGATGAGGATGAAGAG 2449

RESULT 2

ABX71307

ID ABX71307 standard; cDNA; 3664 BP.

XX

AC ABX71307;

XX

DT 14-APR-2003 (first entry)

XX DE Human transcription factor cDNA from clone DKFZphut1_18c19.
XX KW Human; gene; gene therapy; vaccine; disease treatment: detection; ss.
XX OS Homo sapiens.
XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-1B01496.
XX PR 18-AUG-1999; 99US-0149499.
XX PR 28-SEP-1999; 99US-0156503.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX WPI; 2001-327840/34.
XX DR P-PSDB; ABUS2869.
XX CC Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.
XX PS Claim 1; Page 449-450; 1095pp; English.
XX CC This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence encodes a polypeptide described in the disclosure of the invention.
XX SQ Sequence 3664 BP; 1210 A; 734 C; 809 G; 911 T; 0 other;

Alignment Scores:

Pred. No.: 1 68e-272 Length: 3664
Score: 327.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0

US-09-890-549-4 (1-759) x ABX71307 (1-3664)

QY 1 MetGluSerSerProPheAsnArgGlnTTPThrSerLeuSerLeuArgValThrAla 20
Db 94 ATGGATCATCTCCATTTAATAGCGGCATAGAGCTCCTCATCATTTAGGGTAACAGCC 153
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerAlaIleValGluIlePheSerLys 40
Db 154 AAAGAACTTCTCTGTCAACAAGAACAGTATCGCTATTGTGGAATATATCTCCAG 213
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 214 TACCAGAAAGCAGCTGAAAGAAACAAACATGGAGAGAGAGAGTAACACCAGAAATCTC 273
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 274 TCCAGACTTTTAGAAGGGGACCTGACTGTGTTAAAGAGAGAGTGGAGAACCCAGGG 333
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 334 CTGGGAGCAGAGTCTCTACACAGACTCTCTAGGACACAGCAGCAGTATAGGCACAGA 393
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120

Db 394 GCAGACCATCTCTGCTGAAGTGACAGCCACGCTGCTTCTGGAGCCAAAGCTGACCAA 453
 Qy 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
 Db 454 GAAGAAACAAATCCACCCACAGATCTAGACTCAGGTCACTCTCTGAAGCCCTCGTTCAGGCT 513
 Qy 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
 Db 514 CGATATCCCCACATCAAGACCGGTGAGGATCTTTAAAGACCACTCAACAGAAAAGTAAAAA 573
 Qy 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
 Db 574 ATGGAATTTGCTAGGAGATCCAGGCATGAAGTAGAAAAATCAGAAATCAGTGAATAC 633
 Qy 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
 Db 634 ACAGATGCTTCGGGCAAAATAGAGAAATATATGTTCCGCTGAACAGGCTTAAGATGATG 693
 Qy 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
 Db 694 TTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAAGTCAAGT 753
 Qy 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
 Db 754 GGNAGGAAGATCTCTGAACACAGCTATTCTTAGATGACTGTGAATAGCCCGAGGTGAG 813
 Qy 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
 Db 814 TTGTGATCTCTACATTTGACTCGGAGAAAAATGACAGTAGACGAAATCTCGAACTTCCA 873
 Qy 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
 Db 874 CGCCTCTCAAAACCTCTATAAGAGTCAATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAAT 933
 Qy 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
 Db 934 CAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAAT 993
 Qy 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
 Db 994 CATAAATGGAGCAAAAGCAGAAATGTCGCCCGAGTCTCTGAGGTCTGATCACTCCCATCAG 1053
 Qy 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
 Db 1054 GAAGGGGAAAGATTCTGCAAAATGAGATAGCTGGCAGTCCGTTCCACCCCTCCCGAA 1113
 Qy 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
 Db 1114 GATGACTCCCGTACTCCCGAGGTAAAGTAGAGTGAAGTTCACACGCTGTCCATCCCAAGCCA 1173
 Qy 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
 Db 1174 CTAAGTCCAGATTCAGAGCTCCAGTCTTTCTGAAAGTTCCTCCCAAGCAATGAAG 1233
 Qy 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
 Db 1234 AAGTTTCAGGCACCTGCAAGACAGACCTGCGTGGAAATGTCAGAACACAGTCTATCCAATG 1293
 Qy 401 GluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
 Db 1294 GAGCGTCTCTCGCCCAACACAGCAGGTGTTTCAATCAGCTGCTTCGTTGCTCTCTATTGC 1353
 Qy 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
 Db 1354 AACACAAACTCAGTCTAGAACATATGTCATCTTTACATGAAGAAATCTATTGTAAGCT 1413
 Qy 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
 Db 1414 CACTTCAATCAACTCTTTAAATCTAAGGGCACTATGATGAAGCTTTGGCCACAGACCA 1473
 Qy 461 HisLysAspLeuTyrAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAlaGln 480
 Db 1474 CACAAGGATCTATGGGCAAGCAAAATGAAACCGAAGAGATTTTGGAGAGACCCAGCCAG 1533

Qy 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
 Db 1534 CTTGCAATGCAAGGAGAGACCCCTCACAGCCAGGGTGAAGATGCCCTATTGCTTAAG 1593
 Qy 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
 Db 1594 GTGGGTGCTCTGGTGCAGATATGGAAGCCCAAGGCTCTCTCAGCAGGAGGAAGAC 1653
 Qy 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
 Db 1654 AAGCAGCTGAAACCAAGAAAGCTGAGGATCGCTGGCCACCCCTCACTGAACCTTGAAGT 1713
 Qy 541 SerGlySerAlaLeuGluGlyLysMetSerLysProLysTyrProProGluAsp 560
 Db 1714 TCAGGAAGTGCCTTGGAGGAAGGATCAAAATGTCAAGGCCAAATGGCTCTCTGAAGAC 1773
 Qy 561 GluLysSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArg 580
 Db 1774 GAAATCAGCAAGCCGAAATTCCTGAGGATGTCTAGATCTGAGAAAGCTTCAAGACCA 1833
 Qy 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
 Db 1834 TCTTCTTCACTGAAGGAAAGAGCCGCCATTCCTGTAGCAGCTTCAATTCGAAGCACC 1893
 Qy 601 SerValLysSerProLysThrValSerProProIleArgLysGlyTyrSerMetSerGlu 620
 Db 1894 TCTGTCAAGACCCCAAAATCTGTCTCCCACTATCAGGAAGGCTGGAGCATGTCAAG 1953
 Qy 621 GlnSerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
 Db 1954 CAGAGTCAAGAGTCTGTGGTGGAGAGAGTTCGAGAAAGGAAACAAGTGGAAATGCCAAG 2013
 Qy 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
 Db 2014 GCTTCTAAGAAAGATGGGAATGTGGGAAAAACAACCTGGCAAAACAAGAAATCTAAAGGA 2073
 Qy 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
 Db 2074 GAGACGGGAAGAGAAAGTAAGGAAGGTATAGTTTGGAGATGGAGATGAGAAATCTTGTA 2133
 Qy 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGln 700
 Db 2134 GAAATGTCGACACTCCGATGAAGATGATACAGCTTCTCTCAACAACTATCCACAA 2193
 Qy 701 GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPheThr 720
 Db 2194 GAACCCAAAGTCTCTGAATTCGTTCGAGTTTGTAGACACACCTTTGCTGAAGAAATCACT 2253
 Qy 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu 740
 Db 2254 ACTCAGAATCAGAAATCCCAAGGATGTGGAATCTCGGAGGGAGAAAGTGTCAAAGAGCTC 2313
 Qy 741 SerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
 Db 2314 TCTGTGGAAAGACAGATAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2370
 RESULT 3
 AAA53826
 ID AAA53826 standard; DNA; 3705 BP.
 XX
 AC AAA53826;
 XX
 DT 03-JAN-2001 (first entry)
 XX
 DE Sequence encoding lipid associated protein (LIPAP) 2766980CB1.
 XX
 KW Lipid associated protein; LIPAP; treatment; prophylaxis; agonist;
 KW antagonist; antibody; cardiovascular disease; neurological disease;
 KW gastrointestinal disease; lipid metabolism; detection;
 KW amplification; monitoring; hybridisation; antisense; triplex;
 KW ribozyme; screening; immunoassay; ds.
 XX

OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 137..2416
FT /*tag= a
FT /product= Lipid associated protein
XX
PN WO200049043-A2.
XX
PD 24-AUG-2000.
XX
PF 18-FEB-2000; 2000WO-US04160.
XX
PR 19-FEB-1999; 99US-0120703.
PR 08-JUL-1999; 99US-0142762.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Hillman JL, Yue H, Azimzai Y, Baughn MR, Tran B;
XX
DR WPI: 2000-549264/50.
DR P-PSDB; AAY97286.
XX
XX New human lipid-associated proteins, nucleic acids, and antibodies,
PT useful for diagnosis, treatment and prevention of e.g. cardiovascular
PT disease
XX
PS Claim 4; Page 87-88; 93pp; English.
XX
CC Lipid-associated proteins (LIPAP) can be used for treating or
CC preventing disorders associated with decreased expression of LIPAP,
CC for screening for agonists or antagonists of LIPAP, and to raise
CC specific antibodies. Antagonists and agonists of LIPAP are useful
CC for treating diseases associated with reduced or increased levels of
CC LIPAP, e.g. cardiovascular, neurological and gastrointestinal
CC diseases and disorders of lipid metabolism. Fragments of the nucleic
CC acid encoding LIPAP are useful for detection of full length coding
CC sequences, in hybridization and/or amplification assays or for
CC diagnosis or monitoring. Nucleotides encoding LIPAP are used
CC to screen for compounds that specifically modify LIPAP expression,
CC for recombinant production of LIPAP, in gene therapy, as a source of
CC therapeutic antisense, triplex-forming, or ribozyme agents and for
CC genomic mapping. Antibodies to the proteins are used for diagnosis
CC and monitoring of LIPAP-associated disease by immunoassay, as
CC antagonists, in competitive drug screens and for affinity
CC purification of natural LIPAP.
XX
SQ Sequence 3705 BP; 1217 A; 743 C; 826 G; 919 T; 0 other;

Alignment Scores:
Pred. No.: 1.71e-272 Length: 3705
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-890-549-4 (1-759) x AAAS3826 (1-3705)

QY 1 MetGluSerSerProPheAsnArgGlnTTPThrSerLeuSerLeuArgValThrAla 20
DB 137 ATGGAATCATCTCCATTTAATAGCGCAATGACCTACTATCATTTGAGGGTAACAGCC 196

QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
DB 197 AAAGAACTTCTCTTGTCAACAGAACAAAGTATCGGCTATTGTGGAATATATTCCTCAAG 256

QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
DB 257 TACCAGAAAGCAGCTCAAGAAACAAACATGGAGAGAGAGAGATTAACCCGAAATCTC 316

QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80

317 TCCAGCAGCTTTAGAAAGGGGACCTGACTGTGTTAAAGAAAGAGTGGAGAACCCAGGG 376
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
DB 377 CTGGAGAGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACTGAGATTAGGCACACA 436
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
DB 437 GCAGACCATCTCTCTGCTGTAAGTCAAGCCACGCTGCTCTCTGGAGCCCAAGCTGACCAA 496
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
DB 497 GAAGAACAAATCCACCCAGATCTAGACTCAGTCACTCTCTGGAAGCCCTCGTTTCAAGGT 556
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
DB 557 CGATATCCCCACATCAAGGACGCTGAGGATCTTAAGACCACTCAACAGAAATGTAATAA 616
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
DB 617 ATGGAAATTTGTCTAGGAGATCCAGGATGAAGTAGAAATAATCAGAAATCAGTGAATAAC 676
QY 181 ThrAspAlaSerGlyLysIleGluSerArgHisGluValGluLysSerGluIleSerGlu 200
DB 677 ACAGATGCTTCGGGCAAAATAGAGAAATATATATGTTCCGCTGAACAGGCTTAAGATGATG 736
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
DB 737 TTTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGTCAAGT 796
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
DB 797 GGAAGGAAGATCTCTGAAACACAGCTATTCTCTAGATGACCTGGAAATAGGCCAGGTGAG 856
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
DB 857 TTGTCATCTTACATTTGACTCGGAGAAAAATGAGAGTAGACGAAATCTGGAACTTCCA 916
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
DB 917 CCGCTCTCAGAAACCTCTATAAAGATGCGAATGCCAAGTAGCCAGGACGTGTGTCCAAA 976
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLysLysAlaSerGlyGlyGluIleLysIle 300
DB 977 CAAGCAGCTCAACCAACTATACAAATGAGCTGAAGCCAGTGTGTGGCAAAATCAAAAT 1036
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
DB 1037 CATAAAATGGAGCAAAAGGAGAAATGTGCCCCAGTCTCTGAGGTCTGCATCACCCTCAG 1096
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
DB 1097 GAAGGGGAAAAAGATTCTGCAAAATGAGAATAGCCTGGCAGTCCGTTCCACCCCTGCCGAA 1156
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
DB 1157 GATGACTCCCGTGACTCCCGAGTTAAGAGTAGGTTCAACAGCCCTGTCCATCCCAAGCCA 1216
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
DB 1217 CTAAGTCCAGATTCAGAGGCTCCAGTCTTTCTGAAAGTTCTCTCTCCCAAGCAATGAAG 1276
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
DB 1277 AAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGAATGTCAGAGAGACAGTCTATCCAAATG 1336
QY 401 GluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
DB 1337 GAGGCTCTCTTGCGCAACCCAGCAGGTGTTTTCATCATGAGTGTCTCGTCTCTCTATTGC 1396
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
DB 1397 AACACAAACTCAGTCTAGGAACATATGATCTTTTACATGGAAAGATCTATTGTAAGCCT 1456

QY 441 HisPheAenGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1457 CACITTCATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGCTTTGGGCACAGACCA 1516
QY 461 HisLysAspLeuTtpAlaSerLysAsnGluGluLeuLeuGluArgProAlaGln 480
Db 1517 CACAAGGATCTATGGCCAGCAAAATGAAACCAAGAGATTTTGGAGACCAAGCCAG 1576
QY 481 LeuAlaAenAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1577 CTTGCAATGCAAGGAGACCCCTCACAGCCAGGGGTAGAGATGCCCTATTGCTTAAG 1636
QY 501 ValGlyValLeuAlaAaSerMetGluAlaLysSerGlnGlnGluLysGluAsp 520
Db 1637 GTGGGTGTCCTGGTGCAGATGATGAAGCAAGCCCTCTCTCAGCAGGAGAGCAAGAC 1696
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
Db 1697 AAGCCAGCTGAAACCAAGAGCTGAGGATCCCTGGCCACCCCTCACTGAATGGAAGT 1756
QY 541 SerGlySerAlaLeuGluGluGlyLysMetSerLysProLysTyrProProGluAsp 560
Db 1757 TCAGGAAGTGCCTTGGAGAGAGGATCAAAATGTCAAGGCCCAATGGCTCTCTGAAGAC 1816
QY 561 GluLysSerLysProGluValProGluAspValAspLeuAspLysLysLysLeuArg 580
Db 1817 GAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTGAAGAAGCTTAAGACGA 1876
QY 581 SerSerSerLysGluArgSerArgProPheThrValAlaAaSerPheGlnSerThr 600
Db 1877 TCCTTCCTCAGTGAAGGAAAGCGCCCTTCTCTAGCAGCTTCATTTCAAAGCACC 1936
QY 601 SerValLysSerProLysThrValSerProIleArgLysGlyTyrSerMetSerGlu 620
Db 1937 TCTGTCAGAGCCCAAAATCTGTCCTCCCTATCAGGAAGGCTGAGCATGTTCAGAG 1996
QY 621 GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
Db 1997 CAGAGTGAAGAGTCTGTGGTGGAGAGTTGCAGAAAGGAAACAAGTGGAAATGCCAAG 2056
QY 641 AlaSerLysLysAenGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660
Db 2057 GCITCTAAGAGAAATGGGAATGTGGGAAAAACAACCTGGCAAAACAAGAATCTAAGGA 2116
QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAenGluAenLeuVal 680
Db 2117 GAGACAGGAGAGAGAGTAAGAGAGTCTAGTTTGGAGATGAGATGAGATCTTGA 2176
QY 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGln 700
Db 2177 GAAATGGTCAGACTCCGATGAAGATGATAACAGCTTCTCAACAACATCTCCACA 2236
QY 701 GluProLysSerLeuAenTtpSerSerPheValAspAsnThrPheAlaGluGluPheThr 720
Db 2237 GAACCCCAAGTCTCTGAATGGTGCAGTTTGTAGACAACACCTTGTGAGAAATTCAC 2296
QY 721 ThrGlnAenGlnLysSerGlnAspValGluLeuTrpGluGlyValValLysGluLeu 740
Db 2297 ACTCAGATCAGAAATCCCAAGATGTGGAACTCTGGAGGAGAGATGTGTCAAGAGCTC 2356
QY 741 SerValGluGluGlnLysLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 2357 TCTGTGGAGAACAGATGAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2413

RESULT 4

AAC77143
ID AAC77143 standard; cDNA; 3711 BP.

XX AAC77143;

XX AC AAC77143;

DT 08-FEB-2001 (first entry)

XX

Human ORFX ORF2698 polynucleotide sequence SEQ ID NO:5395.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; anticholesterolemia; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach M;

WPI; 2000-602362/57.

P-PSDB; AAB42934.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease.

Claim 5; Page 4578-4580; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective; osteopathic; anticonvulsant; antirheumatic; immunosuppressant; immunostimulant; cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 3711 BP; 1215 A; 747 C; 830 G; 919 T; 0 other;

Alignment Scores:

Pred. No.: 3,09e-271 Length: 3711
Score: 3909.50 Matches: 758
Percent Similarity: 99.74% Conservative: 0
Best Local Similarity: 99.74% Mismatches: 1
Query Match: 99.55% Indels: 1

DB: 21 Gaps: 1
US-09-890-549-4 (1-759) x AAC77143 (1-3711)
QY 1 MetGluSerSerProPheAsnArgGlnThrThrSerLeuSerLeuArgValThrAla 20
DB 142 ATGAATCATCTCCATTTAATAGACGCCAATGACCTCACTATCATTTGAGGGTAACAGCC 201
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerAlaIleValGluIlePheSerLys 40
DB 202 AAGAAGCTTTCTTGTCAACAGAACACAGTCACTCGGTATTGTGGAATATTCTCCAAG 261
QY 41 TyrGlnLysAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
DB 262 TACCAGAAAGCAGCTCAAGAAACAAACATGGAGAACAGAGAGTAACACCCGAAATCTC 321
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
DB 322 TCCAGACACTTTAAGAAGGGGACCTGACTGTGTAAAGAAAGTGGGAGAACCCAGGG 381
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
DB 382 CTGGAGACAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACTGAGATTAGGCACAGA 441
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
DB 442 GCAGACCATCTCTGCTGAAGTGACAGCCAGCTGCTTCTGGAGCCAAAGCTGACCAA 501
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
DB 502 GAAGAACAAATCACCCAGACTCTAGACTCAGTCACTCTCTGAAAGCCCTCTGTCAGGGT 561
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
DB 562 CGATATCCCCCATCAAGACCGGTGAGGATCTTAAAGACCACTCAACAGAAAGTAAAAA 621
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
DB 622 ATGGAATGTCTAGGAGATCCAGGCATGAAGTAGAATAATCAGAAATCAGTGAATAAC 681
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
DB 682 ACAGATGCTTCGGGCAAAATACAGAAATATAATGTTCCGCTGAACAGGCTTAAGATGATG 741
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
DB 742 TTGGAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAGGTGCAAGT 801
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
DB 802 GGAAGAGATCTCTGAAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCAGGTGAC 861
QY 241 LeuSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
DB 862 TTGTCATCTTACATTTGACTCGAGAAAATAGAGATAGACGAAATCTGGAACTTCCA 921
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaValSerLys 280
DB 922 CGCTCTCAGAAACCTCTATAAGATCGAATGCCAAGTACCGAGTACCGAGCTGTGTCAAA 981
QY 281 GlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
DB 982 CAAGACAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGGCGAAATCAAAAT 1041
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
DB 1042 CATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCTGAGGTCTGCATCACCCATCAG 1101
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
DB 1102 GAAGGGGAAAAGATTTCTGCAATAGAGATAGCTGCACTCGTTCACCCCTGCCGAA 1161
QY 341 AspAspSer---ArgAspSerGlnValLysSerGluValGlnGlnProValHisProLys 359

DB 1162 GNTGACTCCCGAGGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCTGTCCATCCCCAAG 1221
QY 360 ProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMet 379
DB 1222 CCACTAAGTCCAGATTCCAGAGCTCCAGTCTTTCTGAAAGTTCTCTCCCAAGCAATG 1281
QY 380 LysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrPro 399
DB 1282 AAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTCAGAACACAGTCTATCCA 1341
QY 400 MetGluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyr 419
DB 1342 ATGAGCGCTCTCTTGGCCACACAGCAGGTGTTTCAATCAGCTGCTTCGTTGCTCTAT 1401
QY 420 CysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLys 439
DB 1402 TGCAACAACAACCTCAGCTAGGAAACATATGCATCTTTTACATGGAAGAATCTATTGTAAG 1461
QY 440 ProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArg 459
DB 1462 CCTCATTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGGACACAGA 1521
QY 460 ProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAla 479
DB 1522 CCACACAGGATCTATGGCAAGCAAAATGAAACCGAAGAGATTTTGGAGAGACAGCC 1581
QY 480 GlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAla 499
DB 1582 CAGCTTGCATAATGCAAGGGAGACCCCTCACAGGCCAGGGGTAGAAGATGCCCCATTGCT 1641
QY 500 LysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGlu 519
DB 1642 AAGTGGGTGCTCTGGTGCAGATGTAAGAACCCAGGCCCTCTCTCAGCAGGAGAAAGAA 1701
QY 520 AspLysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGly 539
DB 1702 GACAAAGCAGCTGAACCAAGAGAGTGAGGATCGCTGGCCACCCCCCTGAACTTGGGA 1761
QY 540 SerSerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTrpProProGlu 559
DB 1762 AGTTCAGAAAGTCTGGAGAGAGGATCAAAATGTCAAAAGCCCAATGCGCTCTCTGAA 1821
QY 560 AspGluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArg 579
DB 1822 GACAAATCAGCAAGCCCGAAGTTCTGAGGATGTCGATCTAGATCTGAAGAGAGTAAAG 1881
QY 580 ArgSerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSer 599
DB 1882 CGATCTTCTCAGTGAAGGAAAGAGCCGCCCATTCCTGTAGCAGCTTCATTTCAAAGC 1941
QY 600 ThrSerValLysSerProLysThrValSerProProIleArgLysGlyTrpSerMetSer 619
DB 1942 ACCTCTGTCAAGAGCCCAAAAGCTGTGTCGCCACCTATCAGGAAAGGCTGGAGCATGTCA 2001
QY 620 GluGlnSerGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAla 639
DB 2002 GACACAGTCAAGAGTCTGTGGTGGAGAGTTCAGAAAGGAAACAAAGTGGAAATGCC 2061
QY 640 LysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLys 659
DB 2062 AAGGCTTCTAAGAAAGATGGGAATGGGAATGGGAATAACCACTGGCAAAACAAAGAAATCTAA 2121
QY 660 GlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeu 679
DB 2122 GGAGACAGAGGAGAGAGAGTAAAGAGTCAATGTTTGGAGATGGAGATGAGAAATCTTT 2181
QY 680 ValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerPro 699
DB 2182 GTAGAAAATGGTGCAGACTCCGATCAAGATGATAACAGCTTCTCTCAACACAACTCTCCA 2241
QY 700 GlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPhe 719

Db	2242	CAAGAACCCCAAGTCTCTGAATTGTCGAGTGTTCATAGACACACACCTTTGCTGAGAAATTC	2301
Qy	720	ThrThrGlnAsnGlnLysSerGlnAspValGluLeuTyrpGluGlyGluValValLysGlu	739
Db	2302	ACTACTCAGAATCAGAATCCAGAGATGTGGAACCTCTGGAGGGAGAGTGGTCAAGAG	2361
Qy	740	LeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu	759
Db	2362	CTCTCTCTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG	2421
RESULT 5			
ID	AAF55697 standard; DNA; 3650 BP.		
XX	AC	AAF55697;	
XX	DT	11-JUN-2001 (first entry)	
XX	DE	Human EPLIN (epithelial protein lost in neoplasm)-beta isoform DNA.	
XX	KW	Human; EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;	
XX	KW	EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;	
XX	OS	gene therapy; cancer; ss.	
XX	OS	Homo sapiens.	
XX	Key	Location/Qualifiers	
FT	CDS	102..2384	
FT	FT	/*tag= a	
FT	FT	/transl_except= "(pos: 933..935, aa: Val)"	
FT	FT	/transl_except= "(pos: 1131..1136, aa: Arg)"	
FT	FT	/transl_except= "(pos: 1587..1589, aa: Asp)"	
FT	FT	/transl_except= "(pos: 1635..1637, aa: Lys)"	
FT	FT	/transl_except= "(pos: 2068..2070, aa: Ser)"	
FT	FT	/transl_except= "(pos: 2076..2078, aa: Ser)"	
FT	FT	/transl_except= "(pos: 2139..2141, aa: Ile)"	
FT	FT	/transl_except= "(pos: 2196..2198, aa: Ser)"	
FT	FT	/transl_except= "(pos: 2214..2216, aa: Ser)"	
FT	FT	/product= "EPLIN (epithelial protein lost in neoplasm)"	
XX	XX	WO200118019-A1.	
XX	XX	15-MAR-2001.	
XX	XX	08-SEP-2000; 2000WO-US24689.	
XX	XX	08-SEP-1999; 99US-0153024.	
XX	PA	(REGC) UNIV CALIFORNIA.	
XX	PI	Chang DD, Maul RS;	
XX	XX	WPI; 2001-244555/25.	
XX	DR	P-PSDB; AAB67701.	
XX	XX	New tumor suppressor protein EPLIN, useful as a marker for diagnostic,	
XX	PT	prognostic and therapeutic applications over the course of cell	
XX	PT	proliferative disorders associated with EPLIN -	
XX	PS	Claim 4; Page 43-44; 59pp; English.	
XX	XX	The present sequence encodes a human EPLIN (epithelial protein lost in	
XX	CC	neoplasm)-beta isoform. The specification also describes EPLIN-alpha.	
XX	CC	EPLIN is a tumour suppressor protein, whose expression is altered in	
XX	CC	multiple common human tumour types. EPLIN nucleic acids and proteins are	
XX	CC	used in screening assays to detect molecules that specifically bind to	
XX	CC	EPLIN nucleic acids, proteins or derivatives and thus have potential use	
XX	CC	as agonist or antagonist of EPLIN, in particular molecules that affect	
XX	CC	cell proliferation. Thus the assays are useful for screening molecules	
XX	CC	with potential utility as anticancer drugs or lead compounds for drug	
XX	CC	development. EPLIN nucleic acids, proteins are useful for detecting a	
XX	CC	cell proliferative disorder in a subject. EPLIN polynucleotides are	
XX	CC	useful in gene therapy techniques. EPLIN is useful as a marker that	

942 CAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAAT 1001
QY HisLysMetGluGlnLysGluAsnValProGlyProGluValCysIleThrHisGln 320
Db CATAAATGAGGCAAAAGGAGATGTCCCGCCAGGTCCTGAGTCTGCATCAACCCATCAG 1061
QY GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db GAAGGGAAGAAGATTCTCGCAATGAGATAGCTGGCAGTCCGTTCCACCCCTGCCGAA 1121
QY AspAspSer---ArgAspSerGlnValLysSerGluValGlnGlnProValHisProLys 359
Db GATGACTCCCGAGTGAAGTCCAGGTTAAGAGTGAGGTTCAACAGCGTGTCCATCCCAAG 1181
QY ProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMet 379
Db CCCTAAGTCCAGATCCAGAGGCTCCAGGCTTCCTGAAAGTTCCTCTCCCAAGCAATG 1241
QY LysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrPro 399
Db AAGAGTTTCAGGCACCTGCAAGAGACCTGGTGGATGTGAGAGACAGTCTATCCA 1301
QY MetGluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyr 419
Db ATGAGCGTCTCTGGCCAAACAGCAGGCTGTTTACATCAGCTGCTTCGTTGCTCTAT 1361
QY CysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLys 439
Db TGCACCAACAACTCATGCTPAGGAACATATGCTATTTACATGGAAGATCTATTGTAAG 1421
QY ProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArg 459
Db CCTCCTCAATCACTCTTTAAATCTAAGGGCACTATGATGAAGGCTTGGGCACAGA 1481
QY ProHisLysAspLeuTrpAlaSerLysAsnGluGluGluIleLeuGluArgProAla 479
Db CCACACAGGATCTATGGCAACCAAAATGAAACGAGAGATTTTGAGAGACAGCC 1541
QY GlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAla 499
Db CAGCTTGCAAAATCAAGGAGACCCCTCACAGCCAGGGGTAGAAAATGCCCTATGCT 1601
QY LysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysGlu 519
Db AAGTGGGTGCTCTGGCTGCAAGTATGGAAGCCAGGCGCTCTCTCAGCAGGAGAGAA 1661
QY AspLysProAlaGluThrLysLysLeuArgIleAlaTrpProProProThrGluLeuGly 539
Db GACAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGGCCCTCCCGCCCTGAACTTGA 1721
QY SerSerGlySerAlaLeuGluGluIleLysMetSerLysProLysTrpProGlu 559
Db AGTTCAAGGAAGTCCCTTGAGGAAGGATCAAAATGTCAAAGCCCAATGGCTCTGAA 1781
QY AspGluIleSerLysProGluValProGluAspValAspLeuLysLysLeuArg 579
Db GAGCAATCAGCAAGCCCGAAGTTCCTGAGATGTGATCTAGATCTGAGGAAGCTAAGA 1841
QY ArgSerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSer 599
Db CGATCTTCTTCACTGAGGAAGAAGCCCGCCATTCACGTGACAGCTTCATTTCAAGC 1901
QY ThrSerValLysSerProLysThrValSerProIleArgLysGlyTrpSerMetSer 619
Db ACTCTGTCAAGAGCCCAAAACTGTGTCACCTATCAGGAAGGCTGGAGCATGCA 1961
QY GluGlnSerGluLysValGlyArgValAlaGluArgLysGlnValGluAsnAla 639
Db GAGCAGATGAAGATCTCTGGGTGGAAGATGCAAGAAAGAAACAAGTGGANAATGCC 2021
QY LysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLys 659
Db AAGGCTTTCAAGAGATGGGAATGGGAAAAACAACCTGGCAAAAACAAGAAATTTAA 2081

660 GlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeu 679
Db GGAGAGACAGGCAAGAGAGAGTAAAGCAAGTTCATAGTTTGGAGATGGAGATTTT 2141
QY ValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnClnSerPro 699
Db GTAGAAAATGGTGCAAGTCCGATGAAGATGATAACAGCTTCTCTCAACAACAATTTCCA 2201
QY GlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPhe 719
Db CAAGAACCCAGTTTTTGAATTTGTCGAGTTTGTAGACACACCTTTGCTGGAAGATTC 2261
QY ThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValLysGlu 739
Db ACTACTCAGATCAGAAATCCAGGATGTGAACTTTGGAGGGAGAGTGGTCAAGAG 2321
QY LeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGlu 759
Db CTCTCTGGGAAGACAGATAAGAGAAATCGGTATTATGATGAGGATGAGGATCAAGAG 2381
RESULT 6
AAF55696
ID AAF55696 standard; DNA; 3543 BP.
XX
AC AAF55696;
XX
DT 11-JUN-2001 (first entry)
XX
DE Human EPLIN (epithelial protein lost in neoplasm)-alpha isoform DNA.
XX
KW Human; EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;
KW EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;
KW gene therapy; cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 474..2276
FT /cds= a
FT /transl_except= "(pos: 825..827, aa: Val)"
FT /transl_except= "(pos: 1479..1481, aa: Asp)"
FT /transl_except= "(pos: 1627..1629, aa: Lys)"
FT /transl_except= "(pos: 1860..1861, aa: Ser)"
FT /transl_except= "(pos: 1968..1970, aa: Ser)"
FT /transl_except= "(pos: 2031..2033, aa: Ile)"
FT /transl_except= "(pos: 2088..2090, aa: Ser)"
FT /transl_except= "(pos: 2106..2108, aa: Ser)"
FT /product= "EPLIN (epithelial protein lost in neoplasm)"
XX
WO200118019-A1.
XX
XX 15-MAR-2001.
XX
XX 08-SEP-2000; 2000WO-US24689.
XX
XX 08-SEP-1999; 99US-0153024.
XX (REGC) UNIV CALIFORNIA.
XX Chang DD, Maul RS;
XX
XX WPI; 2001-244555/25.
XX P-PSDB; AAB67700.
XX
XX New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
XX prognostic and therapeutic applications over the course of cell
XX proliferative disorders associated with EPLIN -
XX
XX Claim 4; Page 42-43; 59pp; English.
XX
XX The present sequence encodes a human EPLIN (epithelial protein lost in
XX neoplasm)-alpha isoform. The specification also describes EPLIN-beta.
XX

CC EPLIN is a tumour suppressor protein, whose expression is altered in
 CC multiple common human tumour types. EPLIN nucleic acids and proteins are
 CC used in screening assays to detect molecules that specifically bind to
 CC EPLIN nucleic acids, proteins or derivatives and thus have potential use
 CC as agonist or antagonist of EPLIN, in particular molecules that affect
 CC cell proliferation. Thus the assays are useful for screening molecules
 CC with potential utility as anticancer drugs or lead compounds for drug
 CC development. EPLIN nucleic acids, proteins are useful for detecting a
 CC cell proliferative disorder in a subject. EPLIN polynucleotides are
 CC useful in gene therapy techniques. EPLIN is useful as a marker that
 CC can be diagnostically, prognostically and therapeutically used over
 CC the course of a cell proliferative disorder associated with EPLIN.
 XX

SQ Sequence 3543 BP; 1152 A; 711 C; 771 G; 909 T; 0 other;

Alignment Scores:

Pred. No.:	2,866-249	Length:	3543
Score:	3603.50	Matches:	695
Percent Similarity:	99.01%	Conservative:	3
Best Local Similarity:	98.58%	Mismatches:	1
Query Match:	91.76%	Indels:	6
DB:	22	Gaps:	1

US-09-890-549-4 (1-759) x AAF55696 (1-3543)

QY	56	AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys	75
DB	159	AACACCGAAATCTCTCCACACCTTTTAAAGAGGGGACCCCTGACTGTGTTAAAGAGAAG	218
QY	76	TrpGluAsnProGlyLeuGluValAlaGluSerHisThrAspSerLeuArgAsnSerSerThr	95
DB	219	TGGAGAAACCCAGGGCTGGAGCAGAGTCTCACACAGACTCTTACGGAAACAGCAGCACT	278
QY	96	GluLeuArgHisArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGly	115
DB	279	GAGATTAGGCACAGACAGACCATCTCTCTGTAAGTGACAAACCCAGCGTCTTCTGGA	338
QY	116	AlaLysAlaAspGlnGluGluLeuHisProArgSerArgLeuArgSerProProGlu	135
DB	339	GCCAAAGCTGACCAAGAGAACAATCCACCCAGATCTAGACTCAGTCCACCTCTCTGAA	398
QY	136	AlaLeuValGlnGlyArgTyrProHisLysAspGlyGluAspLeuLysAspHisSer	155
DB	399	GCCTCTGTTACAGGTGCGATATCCCCACATCAAGACCGGTAGGATCTTAAAGACCACTCA	458
QY	156	ThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSer	175
DB	459	ACAGAAAGCTAATAAATGGAAAAATTGTCTAGAGAAATCCAGGCATGAAGTAGAAAAATCA	518
QY	176	GluLeuSerGluAsnThrAspAlaSerGlyLysLysLeuLysTyrAsnValProLeuAsn	195
DB	519	GAATTCAGTGAACACACAGATGCTTCGGCAAAATAGAGAAATATAATGTTCCGCTCAAC	578
QY	196	ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysLysLeuArgAlaGln	215
DB	579	AGGCTTAAGATGATGTTTGAAGAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAA	638
QY	216	SerArgSerAlaSerGlyArgLysLysSerGluAsnSerTyrSerLeuAspAspLeuGlu	235
DB	639	AGCCGAAGTGAAGTGAAGAAAGATCTCTGAAAACAGCTATTCTTAGATGACCTGGAA	698
QY	236	IleGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArg	255
DB	699	ATAGGCCAGGTGAGTTGTCATCTTCTACATTTGACTCGGAGAAAAATGAGAGTAGACGA	758
QY	256	AsnLeuGluLeuProArgLeuSerGluThrSerLysLysAspArgMetAlaLysTyrGln	275
DB	759	AATCTGAAGACTTCACGCCCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCAG	818
QY	276	AlaAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly	295
DB	819	GCAGCTCGCTCCAAACAAACAGACGCTCAACCACTATTACAAATGAGCTGAAGCCAGTGGT	878

QY	296	GlyGluLeuLysLysLysHisLysMetGluGlnLysGluAsnValProProGlyProGluVal	315
DB	879	GGCGAAATCAAAATTCATAAAATGGAGCAAAAGAGATGTCCTCCCCAGGTCTCTGAGGTC	938
QY	316	CysIleThrHisGlnGluGlyGluLysLysSerAlaAsnGluAsnSerLeuAlaValArg	335
DB	939	TGCATCACCCTCAGGAAGGGGAAAGATTTCGCAATGAGAATAGCTTGGCAGTCCGT	998
QY	336	SerThrProAlaGluAspAspSer---A-GaspSerGlnValLysSerGluValGlnGln	354
DB	999	TCCACCCCTGCCGAAGATCACTCCCGAGTGACTCCACAGTTAAGAGTAGGTTCAACAG	1058
QY	355	ProValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSer	374
DB	1059	CCTGTCTCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCTCCAGTCTTTCTGAAAGTTCT	1118
QY	375	ProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGln	394
DB	1119	CTTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTACG	1178
QY	395	LysThrValTyrProMetGluArgLeuAlaAsnGlnValPheHisLysSerCys	414
DB	1179	AAGACAGTCTATCCAAATGGAGCGTCTCTTGGCCAACCCAGCAGGTGTTTCCATCTCAGCTGC	1238
QY	415	PheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGly	434
DB	1239	TTCCGTTGCTCTTATTGCAACCAACAACTCAGTCTAGGAACATATGCACTCTTTACATGGA	1298
QY	435	ArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGlu	454
DB	1299	AGAATCTATGTGAAGCTCCTCAATCAACTCTTTAAATCTAAGGGCACTATGATGAA	1358
QY	455	GlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLysAsnGluAsnGluGluLeu	474
DB	1359	GGCTTTGGGCACAGACCCACACAGGATCTATGGCAAGCAAAATGAAACCAAGAGATT	1418
QY	475	LeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGlu	494
DB	1419	TTGGAGAGACCCAGCCAGCTTCCAAATGCAAGGGAGACCCCTCAGCCAGGGGTAGAA	1478
QY	495	AspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSer	514
DB	1479	AATGCCCTTATTGCTAAGGTGGTGTCTCTGGTGTCAAGTATGGAAGCCAGGCGCTCTCT	1538
QY	515	GlnGlnGluLysGluAspLysProAlaGluThrLysLysLeuArgLysAlaTyrProPro	534
DB	1539	CAGCAGGAGAGAGAACAGACAGCCAGCTGAAACCAAGAAAGCTGAGGATCGCTGGCCACC	1598
QY	535	ProThrGluLeuGlySerSerGlySerAlaLeuGluGlyLysLysMetSerLysPro	554
DB	1599	CCCACTGAACCTTGGAAAGTTTCAGAAAGTCCCTTGGAGGAAGGATCAAAATGTCAAAGCCC	1658
QY	555	LysTyrProProGluAspGluLysSerLysProGluValProGluAspValAspLeuAsp	574
DB	1659	AAATGGCTCTCTGAAGAGCAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCTGATCTAGAT	1718
QY	575	LeuLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPheThrValAla	594
DB	1719	CTGAAGAAGCTAAGACGATCTTCTTCACTGAAGGAAGAAAGCCGCCATTCACTGTAGCA	1778
QY	595	AlaSerPheGlnSerThrSerValLysSerProLysThrValSerProProlleArgLys	614
DB	1779	GCTTCTTCAAGCACCTCTGTCAAGAGCCCAAAACTGTGTCCCCACCTATCAGGAAA	1838
QY	615	GlyTyrSerMetSerGluGlnSerGluGlyValGlyArgValAlaGluArgLys	634
DB	1839	GGCTCGACATGTTCAGACAGAAATGAAGANTCTGTGGGTGGAAGAGTTGCAAGAAAGGAAA	1898
QY	635	GlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGln	654
DB	1899	CAAGTGGAAAAATGCCAAGGCTTCTAAGAGAAATGCGAATGTGGGAAAAACAACCTGGCAA	1958
QY	655	AsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMet	674

Db 1959 AACAAAGAAATTTAAAGGAGAGACAGGAGAGAGTAAGGAGGTCTATGTTGGAGATG 2018
Qy 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeu 694
Db 2019 GAGAAATGAGAAATTTGTAGAAAATGGTGCAGACTCCGATCAAGATGATAACACTTCCIC 2078
Qy 695 LysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThr 714
Db 2079 AAACAACAATTTCCACAAGAACCCAAAGTTTGAATGGTGCAGATTTGTAGACAACACC 2138
Qy 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly 734
Db 2139 TTTGCTGAGCAATTCATCTACTCAGATCAGAAATCCAGGATGTGGAATTTGGGAGGGA 2198
Qy 735 GluValValLysGluLeuSerValGluGluGlnLeuLysArgAsnArgTyrTrpAspGlu 754
Db 2199 GAAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAG 2258
Qy 755 AspGluAspGluGlu 759
Db 2259 GATGAGATGAAGAG 2273

RESULT 7
AAS72746
ID AAS72746 standard; cDNA; 3465 BP.
AC AAS72746;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #8550.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG08559.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID NO 8550; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: the sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3465 BP; 1112 A; 706 C; 773 G; 874 T; 0 other;

Alignment Scores:
Pred. No.: 1.98e-226 Length: 3465
Score: 3285.50 Matches: 688
Percent Similarity: 94.77% Conservative: 1
Best Local Similarity: 94.64% Mismatches: 15
Query Match: 83.66% Indels: 23
DB: 23 Gaps: 5
US-09-890-549-4 (1-759) x AAS72746 (1-3465)
Qy 56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys 75
Db 4 AACACCGAAAAATCTCTCCAGCACTTTAGAAAAGGGACCCCTGACTGTGTAAAGAAGAAG 63
Qy 76 TrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThr 95
Db 64 TGGAGAAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACT 123
Qy 96 GluLeuArgHisArgAlaAspHisProAlaGluValThrSerHisAlaAlaSerGly 115
Db 124 CAGATTAGGCACAGACAGACCATCTCTCTGCTGAGTGACAAAGCCACGCTGCTTCTGGA 183
Qy 116 AlalysAlaAspGlnGluGluGlnHisProArgSerArgLeuArgSerProProGlu 135
Db 184 GCCAAAGCTGACCAAGAAGAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAA 243
Qy 136 AlalysValGlnGlyArgTyrProHisLeuLysAspGlyGluAspLeuLysAspHisSer 155
Db 244 GCCCTCGTTTCAGGGTCGATATATCCCCACATCAAGACGGTGAGGATCTTAAAGACCACTCA 303
Qy 156 ThrGluSerLysLysMetGluAsnCysLeuGlyCysSerArgHisGluValGluLysSer 175
Db 304 ACAGAAAGTAAAAAATGGAAAAATTGCTAGGAGAAATCCAGGCATGAAGTAGAAAAATCA 363
Qy 176 GluLeuSerGluAsnThrAspAlaSerGlyLysLysLeuLysValProLeuAsn 195
Db 364 GAAATCAGTGAAPACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCCTGAAC 423
Qy 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysLysLeuAlaGln 215
Db 424 AGGCTTAAGATGATGTTTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAA 483
Qy 216 SerArgSerAlaSerGlyArgLysLysLeuSerGluAsnSerTyrSerLeuAspLeuGlu 235
Db 484 AGCCGAAGTCAAGTGAAGGAAGATCTCTGAAACACAGCTATTCTCTAGATGACTCTGAA 543
Qy 236 IleGlyProGlyGlnLeuSerSerSerThr-PheAspSerGluLysAsnGluSerArgAr 255
Db 544 ATAGGCCACAGTCAAGTTGTCATCTTTACATTTTGACTCGGAAAAAATGAGAGTAGACG 603
Qy 255 GAsnLeuGluLeuProArgLeuSerGluThrSerLysLysAspArgMetAlalysTyrGl 275
Db 604 AAATCTGGAATCTCCAGCCCTCTCAGAACCTCTATAAAGGATCGAATGCCCAAGTACCA 663
Qy 275 nAlaAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGl 295
Db 664 GGCAGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAATGACTGAATCCAGCGG 723
Qy 295 YGlyGluLysLysLysMetGluGlnLysGluAsnValProProGlyProGluVa 315
Db 724 TGGCAAAATCAAAATTCATAAAACCGGAGGAGAAATGTGCCCCAGGCTCTGAGGT 783

315 lCysIleThrHis-GlnGluGlyGluIysIleSerAlaAsnGluAsnSerLeuAlaVala 335
784 CTGCATCACCCATTGAGGAGGGAAGATTTCTGCAATAGAGATAGCTGCGAGTCC 843
335 lGSerThrProAlaGluAspSer---ArgAspSerGlnValIysSerGluValGlnG 354
844 GTTCCACCCCTGCGAGAGATGACTCCCGAGTGACTCCCGAGTGAAGAGTGAAGTTCAC 903
354 lNProValHisProLysProLysProLysProLysProLysProLysProLysProLys 374
904 AGCTGTGTCATCCCAAGCCTAGTCCAGATTCAGAGCCTCCAGTCTTCTGAAAGTT 963
374 erProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysG 394
964 CTCCTCCCAAGCAATGAAGATTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAGTGC 1023
394 lNlysThrValTy-ProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerC 414
1024 AGAAGACAGTCTATCCCAATGAGAGCTCTCTTGGCCAAACCAAGAGTCTTTCACATCAGCT 1083
414 ysPheArgCysSerTyrcysAsnAsn-LysLeuSerLeuGlyThrTyrrAlaSerLeuHis 433
1084 GCTTCCGTGCTCTATTCGAACCAACAACTCAGTCTAGAACATATGATCTTTACAT 1143
434 GlyArgIleTyrcysLysProHisPheAsnGlnLeuPheLysSer-LysGly-AsnTyrrA 453
1144 GGAAGATCTATTGAGCCTCACTTCAATCACTTTAAATCTTAAGGGCAACTATG 1203
453 spGluGlyPheGlyHisArgProHisLysAspLeu--TrpAlaSerLysAsnGluAsn-G 472
1204 ATGAAGGCTTTGGGCACAGACACACAGGATCTTATTGGGCAAGCAAAATGAAACCG 1263
472 luGluIleLeuGluArgProAla-GlnLeuAlaAsnAlaArgGluThrProHisSerPro 491
1264 AGGAGATTTTGGAGAGACCCAGCCCTGCAATGCAAGGGAGAGCCCTCACGCCCA 1323
492 GlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLys 511
1324 GGGTAGAAGATGCCCTATTGCTAAGTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1383
512 AlaSerSerGlnGlnGluLysGluAspLysProAlaGluThrIlyLysLysLeuArgIleAla 531
1384 GCCTCCTCTCAGCAGGAGAGGAGAGACAGCAGCTGAAACCAAGAGCTGAGGATGCC 1443
532 TrpProProThrGluLeuGlySerGlySerAlaLeuGluGluGlyIleLysMet 551
1444 TGGCCACCCCTCCTGAACTGGAAGTTCAGGAGTGCCTTGGAGGAGGATCAAAATG 1503
552 SerLysProLysTrpProGluAspGluLysSerLysProGluValProGluAspVal 571
1504 TCAGAGCCCAATGGCTCTCTGAGACCAATCAGCAAGCCCAAGTTCTCTGAGGATGC 1563
572 AspLeuAspLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPhe 591
1564 GATCTAGATCTGAAGAAGCTAAGACGATCTTCTACTGAAGGAAGAGCCGCCCATTC 1623
592 ThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPro 611
1624 ACTGTAGCAGCTTCAATTCAAAGCACCTCTGTCAAGAGCCCAAAATCTGTCTCCACCT 1683
612 -IleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValClyArgValAla 631
1684 TATCAGAAAGGCTGGAGCATGTCTAGACAGAGTGAAGAGTCTGTGGTGGAGAGTTC 1743
631 aGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysTh 651
1744 AGAAGGAAACAAGTGGAATGCCAGGCTTCTAAGAGCAATGGGAATGTGGGAAAAAC 1803
651 rThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLys-GluGly-His 670
1804 AACCTGGCAAAACAAGAACTTAAGAGAGAGACAGGAGAGAGTGAAGGAGGTCCAT 1863

671 SerLeuGluMetGluAsnGluAsnLeuVal--GluAsnGlyAlaAspSerAsp---GluA 689
1864 AGTTTCGAGATGAGATGAGATCTTTGTAGAAAATGGTCCAGACTCCCGATTGAAG 1923
689 sPAspAsnSerPhe---LeuLysGlnGlnSer-Pro-GlnGluProLysSerLeu---As 706
1924 ATGATAACCAAGCTTCTCCCAAAACAACATCTTCCACCAAGAACCAAGCTCTCTTGAAT 1983
706 nTrpSer-SerPheValAspAsn---ThrPheAlaGluGluPheThrThr-GlnAsnGln 724
1984 TTGGTCCGAGTTTGTAGACAACCACTTTTCTGAGGAATCTACTACTCCAGAAATCAG 2043
725 LysSerGlnAspValGluLeuTrpGluGlyGluValLysGluLysGluSerValGluGlu 744
2044 AATCCAGAGTGGAACTCTGGAGGGAGAGTGGTCAAGAGCTCTCTGTGGAGAA 2103
745 GlnIleLysArgAsnArgTyrrAspGluAspGluAspGluGlu 759
2104 CAGATAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2148

RESULT 8
ABI98973
ID ABI98973 standard; cDNA; 1754 BP.
XX
AC ABI98973;
XX
DT 21-FEB-2002 (first entry)
XX
DE Human cancer suppressor protein PP624 encoding cDNA.
XX
KW Human; cancer suppressor; disease; cancer; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
444..1349
/*tag= a
/transl_except= (pos:104..106,aa:Xaa)
/transl_except= (pos:561..563,aa:Xaa)
/transl_except= (pos:570..572,aa:Xaa)
/product= "PP624"
/note= "Xaa = unknown"
XX
PN CN1313297-A.
XX
PD 19-SEP-2001.
XX
PF 09-MAR-2000; 2000CN-0111948.
XX
PR 09-MAR-2000; 2000CN-0111948.
XX
PA (SHAN-) SHANGHAI INST ONCOLOGY.
XX
PI Gu J, Yang S;
XX
XX WPI; 2002-042185/06.
DR P-PSDB; ABB56420.
XX
PT Human protein able to suppress growth of cancer cells and its coding
sequence -
XX
PS Claim 5; Page 21-22 Disclosure; 37pp; Chinese.
XX
CC The invention relates to novel human proteins (ABB56417-ABB56425) with
cancer suppressing function, the encoding polynucleotides
CC (ABI98970-ABI98978), the process for preparing the polypeptide, the
CC application of the polypeptide in treating diseases such as cancer, the
CC antagonist of the polypeptide and its medical function and the
CC application of the polynucleotide.
XX
SQ Sequence 1754 BP; 586 A; 377 C; 426 G; 361 T; 4 other;
Alignment Scores:

Pred. No.: 3,23e-183 Length: 1754
 Score: 2679.00 Matches: 523
 Percent Similarity: 98.87% Conservative: 0
 Best Local Similarity: 98.87% Mismatches: 5
 Query Match: 68.22% Indels: 2
 DB: 24 Gaps: 0

US-09-890-549-4 (1-759) x ABI98973 (1-1754)

QY	232	AspAspLeuGluLeuGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLysAsn	251
DB	2	GATGACCTGGAAATAGCCCGAGGTGAGTGTGTCATCTTACATTTGACCTGGAGAAAAT	61
QY	252	GluSerArgArgAsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMet	271
DB	62	GAGAGTAGACGAATCTGGAACTTCCAGCCCTCTCAGAAACCTCTATAAAGGATCGAATG	121
QY	272	AlaLysTyrGlnAlaAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeu	291
DB	122	GCCAACTACAGGACGCTGTGTCCAAACCAAGCAGCTCAACCACTATACAAATGAGCTG	181
QY	292	LysAlaSerGlyGlyGluLeuLysIleHisLysMetGluGlnLysGluAsnValProPro	311
DB	182	AAAGCCAGTGGTGGCGAAATCAAAATTTCAAAATGGAGCAAAAGGAGAAATGGCCCCA	241
QY	312	GlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSer	331
DB	242	GGTCTGTAGGTCTGCATCACTCCAGGAGGAGGAAAGATTTCTGCAATGAGAATAGC	301
QY	332	LeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnValLysSerGlu	351
DB	302	CTGGCAGTCCGTTCACCCCTGCCGAGATGACTCCCGTGACTCCGAGTTAAGAGTGAG	361
QY	352	ValGlnGlnProValHisProLysProLeuSerProLeuSerProAsp-SerArgAlaSerSerLeuSe	371
DB	362	GTTCAACAGCCCTGTCATCCCAAGCCACTAAGTCCAGATCTCCAGAGCTCCAGTCTTTC	421
QY	371	rgLusSerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysVa	391
DB	422	TGAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGT	481
QY	391	lGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGlnValPheHi	411
DB	482	GGAAATGTGAGACACAGCTATATCAATGGAGCGTCTCTGGCCCAACCCAGCAGGTGTTCA	541
QY	411	stleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSe	431
DB	542	CATCAGCTGCTTCGCTGCTCTATTGCAKCAACAACTCAGCTAGGAAACATATGCATC	601
QY	431	rlouHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAs	451
DB	602	TTTACATGGAAGAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCTAAGGCAA	661
QY	451	nTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLysAsnGluAs	471
DB	662	CTATGATGAAGCTTTGGGCACAGACCAACAAAGGATCTATGGCAAGCAAAAATGAAA	721
QY	471	nGluGluLeuLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerPr	491
DB	722	CGAAGAGATTTTGAGAGACACGAGCCAGCTTGCAAAATGCAAGGGAGACCCCTCACGCC	781
QY	491	oGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLy	511
DB	782	AGGGGTAGAAGATGCCCCCTATTGTAAGTGGGTGCTCTGCTGCTCAAGTATGGNAGCAA	841
QY	511	sAlaSerSerGlnGlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAl	531
DB	842	GGCTCTCTCAGCAGGAGAGGAAGCAAGCCAGCTGAAACCAAGAAAGCTGAGGATCGC	901
QY	531	aTrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGlyLysLysMe	551
DB	902	CTGCCCAACCCCTCACTGAACCTGGAAGTTGAGGAAGTGCCTTGAGGAGAGGGATCAAAAT	961

QY	551	tSerLysProLysTrpProGluAspGluIleSerLysProGluValProGluAspVa	571
DB	962	GTCAAAGCCCAATGGCTCTCTGAAGACGAATCAGCAAGCCCAAGTTCCTGAGGATGT	1021
QY	571	lasPleuAspLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPh	591
DB	1022	CGATCTAGATCTGAAGAAAGCTAAGACGATCTTCTCACTGAAGGAAAGAGCGCCCAT	1081
QY	591	eThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPr	611
DB	1082	CACGTGACGACTTCATTTCAAGCACCTCTGTCAAGAGCCCAAAAACCTGTGTCCCCACC	1141
QY	611	oileArgLysGlyTrpSerMetSerGluGlnSerGluSerValGlyGlyArgValAla	631
DB	1142	TATCAGGAAAGCGCTGGAGCATGTCTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAAGTTGC	1201
QY	631	aGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysTh	651
DB	1202	AGAAAGGAAACAAAGTGGAAAATGCAAGGCTTCTTAAGAAAGAAATGCGAATGTGGGAAAAC	1261
QY	651	rThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSe	671
DB	1262	AACCTGGCAAAACAAAGAAATCTAAAGGAGAGCA-GGGAAAGAGAAATGAAGAGGTCATAG	1320
QY	671	rLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsPaS	691
DB	1321	TTTGAGATGCGAATGAGAAATCTTGTAGAAAATGCTGCAGATCCCATGAAGATGATAA	1380
QY	691	nSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheVa	711
DB	1381	CAGCTTCTCAAAACAAATCTCCACAAGAACCCCAAGTCTCTGAATTTGCTCGATTTTGT	1440
QY	711	lasPAsnThrPheAlaGluPheThrGlnAsnGlnLysSerGlnAspValGluLe	731
DB	1441	AGAACACACCTTTGCTGAAGAATTCATCTACTCAGATCAGAAATCCACAGATGTGGAAC	1500
QY	731	uTrpGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArgAsnArgTy	751
DB	1501	CTGGAGGGAGAGTGTGTAAGAGCTCTCTGTGGAAGAAACAGATAAAGAGAAATCGGTA	1560
QY	751	rTyrAspGluAspGluAspGluGlu	759
DB	1561	TTATGATGAGGATGAGATGAAGAG	1585
RESULT 9			
ID	AAI59955	standard; cDNA; 1713 BP.	
XX	XX	AAI59955;	
AC	XX		
DT	XX	22-OCT-2001 (first entry)	
DE	XX	Human polynucleotide SEQ ID NO 3944.	
XX	XX	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;	
KW	XX	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	XX	anyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	XX	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
XX	XX	leukaemia; ss.	
OS	XX	Homo sapiens.	
PN	XX	WO200153312-A1.	
PD	XX	26-JUL-2001.	
XX	XX		
PF	XX	26-DEC-2000; 2000MO-US34263.	
XX	XX		
PR	XX	21-JAN-2000; 2000US-0488725.	
PR	XX	25-APR-2000; 2000US-0552317.	
PR	XX	09-JUL-2000; 2000US-0598042.	
PR	XX	19-JUL-2000; 2000US-0620312.	

PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue A, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-P5DB; AAM40799.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Claim 1; SEQ ID NO 3944; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 1713 BP; 549 A; 401 C; 417 G; 346 T; 0 other;
 Alignment Scores:
 Pred. No.: 5,03e-173 Length: 1713
 Score: 2537.00 Matches: 503
 Percent Similarity: 95.87% Conservative: 8
 Best Local Similarity: 94.37% Mismatches: 20
 Query Match: 64.60% Indels: 4
 DB: 22 Gaps: 0
 US-09-890-549-4 (1-759) x AAI59955 (1-1713)
 QY 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
 Db 121 ATGGAATCATCTCCATTATATAGACGGCAATGGACCTCACTATCATTGAGGGTAACAGCC 180
 QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
 Db 181 AAGAACAATTCTCTTGTCAACAAGAACAAGTCAATCGGCTATTGTGGAATAATTTCTCCAAG 240
 QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
 Db 241 TACCAGAAAGCAGCTGAAGAAACAAACATGGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
 Db 301 TCCACGACCTTTAGAAAGGGGACCTGACTGTGTATAAGAGAAAGTGGGAGAACCCAGGG 360
 QY 81 LeuGluValAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
 Db 361 CTGGGAGCAGAGTCTCACACAGACTCTTACCGAAACAGCAGCAGCAGTGAAGTAGGCACAGA 420
 QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyValLysAlaAspGln 120
 Db 421 GCAGACCATCTCTGCTGAAGTGACACAGCCAGCTGCTTCTGGAGCCAAAGCTGACCAA 480
 QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140

Db 481 GAAGAAACAAATCCACCCAGATCTAGACTCAGGCACCTCTCTGAAGCCCTCGTTCAGGGT 540
 QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
 Db 541 CGATATCCCCCACCATCAAGGACGGTGAGGATCTTAAAGACCCTCTCAACAGAAAGTAAAAA 600
 QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
 Db 601 ATGGAAAAATTGTCTAGGAGATCCAGGCATGAAGTAGAAAAATCAGAGATCAGTGAAC 660
 QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
 Db 661 ACAGATGCTTCGGCAAAATAGAAAAATATATGTTCCGCTGAACAGGCTTAAAGATGATG 720
 QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
 Db 721 TTTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGTCAAGT 780
 QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
 Db 781 GGAAGGAAGATCTCTGAAACACAGCTATTCTCTAGATGACCTGGAAATAGGCCAGGTGAG 840
 QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgAtgAsnLeuGluLeuPro 260
 Db 841 TTTGTCATCTTCTACATTTGACTCCGAGAAAAATAGAGTAGACGAAATTTTGGAACTTCCA 900
 QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
 Db 901 CGCCTCTCAGAAACCTCTATGAAGGATCGAATGGCCAAGTAGCCAGGACGTGTGTCCAAA 960
 QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLysAlaSerGlyGlyGluIleLysIle 300
 Db 961 CAAGACCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGTGGCAAAATCAAAATT 1020
 QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
 Db 1021 CATAAATGGAGCAAAAGGAGAAATGTGCCCCAGTCTCTGAGGTCTGATCACCCTCAG 1080
 QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
 Db 1081 GAAGGGGAAAAAGATTTCGTGCAAAATGAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCAAA 1140
 QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
 Db 1141 GATGACTCCCGTGACTCCCGAGTTAAGAGTGAGGTTCAACAGCCTGTCTCCATCCCAAGCCA 1200
 QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
 Db 1201 CTAAAGTCCAGATTCCAGAGCCTCCAGTCTTCTGAAAAGTTCTCTCCCAAGCAATGAAG 1260
 QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
 Db 1261 AAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGAATGTCAGAAAGACAGCTCTATCCCAATG 1320
 QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
 Db 1321 GAGCGTCTCTTGGCCAAACCAGCAGTGTTCACATCAGTGTCTCCGTTCTCTATTGC 1380
 QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
 Db 1381 AACAAACAACTCAGCTAGGAACATATGCATCTTTATCATGGAAGATCTATTGTAGACCT 1440
 QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
 Db 1441 CACTTCAATCACTTTTAAATCTAAGGGCAACTATGATGAAGGCTTTTGGGAGAGACCACCA 1500
 QY 461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAlaGln 480
 Db 1501 CACAAGGATCTATGGGCAACAAAAATTGAACCGAGGGGTTTGGGAGAGACCACG-CAA 1559
 QY 481 LeuAlaAsnAlaArgGluThrProHis-SerProGlyValGluAsp-AlaProIleAlaL 500

Db 1118 AAGGGCAACTATGATGAAGCTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAA 1177
Qy AsnGluAsnGluGluLeuLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 1178 AATGAAACCAAGAGATTTTGGAGAGACCAGCCAGCTTGCATAATGCAAGGAGAGACCCCT 1237
Qy HisSerProGlyValGluAspAlaProIleAlaIysValGluLeuAlaAlaSerMet 508
Db 1238 CACAGCCAGGGTGAAGATGCCCTTATGCTAAGTGGTGTCTCTGGCTGCAAGTATG 1297
Qy GluAlaLysAlaSerSerGlnGlnGluLysGluAspLysProAlaGluThrLysLysLeu 528
Db 1298 GAAGCCAAAGCCCTCTCTACAGAGGAGAGAGACAGCCAGCTGAACCAAGAGAGCTG 1357
Qy ArgIleAlaTrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548
Db 1358 AGGATCGCTGGCCACCCCTCACTTGAAGTTCAGAAAGTCCCTTGGAGGAGGG 1417
Qy IleIysMetSerLysProLysTrpProProGluAspGluIleSerLysProGluValPro 568
Db 1418 ATCAAAATGTCAAGCCCAATGGCTCTCTGAAGACGAAATCAGCAAGCCGGAAGTTCCT 1477
Qy GluAspValAspLeuAspLeuLysLysLeuArgArgSerSerSerLysLysGluArgSer 588
Db 1478 GAGGATGTCATCTAGATCTGAAGAGCTTAAGACGATCTTCTACTGAAGGAAAGAGC 1537
Qy ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 1538 CGCCCATCTACTAGCAGCTTCAATTCAAAGCACCTCTGTCAAGACCCCAAACTGTG 1597
Qy SerProProIleArgLysGlyTrpSerMetSerSerGluGlnSerGluSerValGlyGly 628
Db 1598 TCCCCACCTATCAGGAAAGCTCGAGCATGTCAAGACAGATGAAGTCTGTGGTGA 1657
Qy ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnVal 648
Db 1658 AGATTTGCAGAAAGGAAACAGTGGAAATGCCAAGCTTCTTAAGAAAGATGGGATGTG 1717
Qy GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 1718 GGAAAAACACCTGGCAAAACAAAGAAATCTAAAGGAGACAGAGGAGAGAAAGTAAAGNA 1777
Qy GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 1778 GGTATAGTTTGGAGATGAGAAATGAGATCTTTAGAAAATGGTGCAGACTCCGATGAA 1837
Qy AspAspAsnSerPheLeuLysGlnGlnSerProGlnGluPheThrGlnAsnGlnLysSerGlnAsp 708
Db 1838 GATGATTAACAGCTTCTCTCAAAACAAACATCTCCACAGAAACCCCAAGTCTCTGAATTGGTGG 1897
Qy SerPheValAspAsnThrPheAlaGluGluPheThrGlnAsnGlnLysSerGlnAsp 728
Db 1898 AGTTTGTAGACACACCTTTGCTGAGAAATTCACCTACTCAGAAATCAGAAATCCCAAGAT 1957
Qy ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArg 748
Db 1958 GTGGAAGTCTGGAGGAGAGAGTGGTCAAGAGCTCTCTGTGGAAGACAGATAAGAGA 2017
Qy AsnArgTyTyAspGluAspGluGlu 759
Db 2018 AATCGGTATTATGATGAGGATGAGGATGAAGAG 2050

RESULT 12

AAZ33566

ID AAZ33566 standard; cDNA; 2783 BP.

XX AAZ33566;

AC AAZ33566;

DT 08-DEC-1999 (first entry)

XX Human breast tumour-associated EST 26.

DE Expressed sequence tag; EST; human; breast; cancer; cytostatic;

XX

KW medicaments; gene therapy; treatment; fat metabolism; ss.

XX Homo sapiens.

XX DE19813835-A1.

XX 23-SEP-1999.

XX 20-MAR-1998; 98DE-1013835.

XX 20-MAR-1998; 98DE-1013835.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;

XX WPI; 1999-528979/45.

XX P-PSDB; AAY48487.

XX Human nucleic acid sequences and protein products from normal breast

XX tissue, useful for breast cancer therapy

XX Claim 1a; 113-114; 206pp; German.

XX This invention describes novel human nucleic acid sequences from normal

XX breast tissue which have cytotostatic activity. The nucleic acid sequences

XX can be used to produce and isolate full-length gene sequences. They can

XX be used to express proteins, which can be used as tools to find an

XX activity against breast cancer. The sequences can be used in sense or

XX antisense form. They are especially useful for medicaments for gene

XX therapy to treat breast cancer and for treating illnesses associated

XX with fat metabolism. AAZ33566-233610 represent expressed sequence tags

XX described in the method of the invention.

XX

SQ Sequence 2783 BP; 901 A; 527 C; 608 G; 747 T; 0 other;

Alignment Scores:

Pred. No.: 7,26e-167 Length: 2783

Score: 2455.00 Matches: 469

Percent Similarity: 99.79% Conservative: 1

Best Local Similarity: 99.58% Mismatches: 1

Query Match: 62.52% Indels: 0

DB: 20 Gaps: 0

US-09-890-549-4 (1-759) x AAZ33566 (1-2783)

Qy 289 AsnGluLeuLysAlaSerGlyGluLeuLysLysLysMetGluGlnLysGluAsn 308

Db 40 AATGAGCTGAAGCAGTGGTGGCAATCAAATTCATAAATGGAGCAAAAGAGAAAT 99

Qy 309 ValProProGlyProGluValCysIleThrHisGlnGluGlyGlyLysIleSerAlaAsn 328

Db 100 GTGCCCCAGGTCCTGAGGTCTGCATCACCCATCAGGAAGGGAAGATTTCTGCAAT 159

Qy 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnVal 348

Db 160 GAGAATAGCTGGCAGTCCCTTCCACCCCTGCCAAGATGACTCCCGTACTCCAGGTT 219

Qy 349 LysSerGluValGlnGlnProValHisProLysProLysProLysProLysProLys 368

Db 220 AAGAGTGGGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 279

Qy 369 SerLeuSerGluSerSerProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388

Db 280 AGTCTTCTGAAGTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAG 339

Qy 389 ThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsnGln 408

Db 340 ACCTGCGTGAATGTCAAGACAGTCTATCCAAATGAGCGTCTCTTGCCCAACAGCAG 399

Qy 409 ValPheHisIleSerCysPheArgCysSerTyCysAsnAsnLysLeuSerLeuGlyThr 428

Db 400 GTGTTTCAATCAGCTGCTTCCGTTCTCTCTATTGCAACAACTCAGTCTAGGAACA 459

QY 429 TyrAlaSerLeuHisGlyArgGlyIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448
Db 460 TATGCATCTTTACATGGAAGAATCTATTGTAAGGCTCACTCAATCAACTCTTTAAATCT 519
QY 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLys 468
Db 520 AAGGGCAACTATGATGAAGCTTTGGGCACAGACCACACAGAGATCTATGGCAAGCAAA 579
QY 469 AsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 580 AATGAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTCAATGCAAGGGAGAGCCCT 639
QY 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMet 508
Db 640 CACAGCCAGGGGTAGAGATGCCCTTATTCCTAAGGGGGTGTCTCTGGCTGCAAGTATG 699
QY 509 GluAlaLysAlaSerSerGlnGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528
Db 700 GAAGCCAAAGGCTCTCTCAGCAGGAGAGAGACAGCCAGCTGAAACCAAGAGCTG 759
QY 529 ArgIleAlaTrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548
Db 760 AGGATCGCTGGCCACCCCACTGAATTTGGAAGTTTCAGGAAGTGCTTTGGAGGAGGG 819
QY 549 IleLysMetSerLysProLysTyrProProGluAspGluLysSerLysProGluValPro 568
Db 820 ATCAAAATGTCAAGCCCAATGCTCTCTGAAAGCAAAATCAGCAAGCCCAAACTGCT 879
QY 569 GluAspValAspLeuAspLeuLysLysLeuArgSerSerSerLysLysGluArgSer 588
Db 880 GAGATGTCATCTAGATCTGAGAGAGCTAAGACGATCTTCTTCACTGAAGGAAAGAGC 939
QY 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 940 CGCCATTCAGTGTAGCAGCTTCAATTTCAAGCACCTCTGTCAAGAGCCCAAACTGTG 999
QY 609 SerProPheArgLysGlyTyrSerMetSerGluGlnSerGluGluSerValGlyGly 628
Db 1000 TCCCCACCTATCAGGAAGGCTGGAGCATGTCAAGCAGAGTCAAGAGTCTGTGGGTGA 1059
QY 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648
Db 1060 AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGGCTTCTAAGAGAGATGGGAATGTG 1119
QY 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 1120 GGAATAACAACCTGGCAAAACAAGATCTTAAGGAGAGACAGGGAAGAAAGTAAGGAA 1179
QY 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 1180 GGTCAATAGTTTGGAGATGGAGATGAGAAATCTTGTAGAAAATGGTCCAGACTCCGATCAA 1239
QY 689 AspAspAsnSerPheLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSer 708
Db 1240 GATGATAACAGCTTCTCTCAAAACAACATCTCAACAAAGAACCCCAAGTCTCTGAATGTCG 1299
QY 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGluAsnGlnLysSerGlnAsp 728
Db 1300 AGTTTTGTAGACACACCTTTGCTGAGAGATTCCTACTACTCAGATCAGAAATCCCAAGAT 1359
QY 729 ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArg 748
Db 1360 GTGGAATCTGGAGGAGGAGAGTGTCAAAGAGCTCTCTGTGGAGAACAGATTAAGAGA 1419
QY 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 1420 AATCGGTATTATGATGAGGATGAGGATGAAGAG 1452

RESULT 13

AAS18588

ID AAS18588 standard; cDNA; 1567 BP.

XX

AC AAS18588;
XX 12-MAR-2002 (first entry)
XX cDNA encoding human sterol regulatory element binding protein 3.
DE XX Human; sterol regulatory element binding protein 3; hSREBP-3;
KW hypothalamus; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
CDS 169..154
FT /*tag= a
FT /product= "Sterol regulatory element binding protein 3,
FT hSREBP-3"
XX CN1309182-A.
XX 22-AUG-2001.
XX 17-FEB-2000; 2000CN-0111698.
XX 17-FEB-2000; 2000CN-0111698.
XX (NATU-) NATURAL HUMAN GENOME NANFANG RES CENT.
XX Li Y, Xu S, Ren S;
XX WPI; 2002-011822/02.
XX P-PSDB; AAU10979.
XX Cholesterol regulatory factor binding protein and its coding sequence -
XX Claim 1; Page 22; 27pp; Chinese.
CC The invention relates to a novel human sterol regulatory element binding
CC protein 3 (hSREBP-3) expressed in human normal hypothalamic tissue and
CC its coding sequence. Also described is the process for preparing the
CC protein and nucleic acid sequence, and the method for detecting hSREBP-3
CC nucleic acid sequence and polypeptides. The present sequence represents
CC the coding sequence of human hSREBP-3 as described in the invention.
XX Sequence 1567 BP; 513 A; 326 C; 391 G; 337 T; 0 other;

Alignment Scores:

Pred. No.: 9.47e-167 Length: 1567
Score: 2449.00 Matches: 468
Percent Similarity: 99.79% Conservative: 2
Best Local Similarity: 99.36% Mismatches: 1
Query Match: 62.36% Indels: 0
Gaps: 24

US-09-890-549-4 (1-759) x AAS18588 (1-1567)

QY 289 AsnGluLeuLysAlaSerGlyGlyGluIleLysIleHisLysMetGluGlnLysGluAsn 308
Db 127 AATGAGCTGAAGCCAGTGTGGCAAAATCAAAATTCATAAAATGGAGCAAAAGAGAAAT 186
QY 309 ValProProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsn 328
Db 187 GTGCCCCCAGGTCTCTGAGTCTGCATCCCATCAGGAGGGGAAAGATTTCTGCAAAAT 246
QY 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnVal 348
Db 247 GAGAATAGCTGGCAGTCCGTTCACCCCTCCGCAAGATGACTCCCGTACTCCAGGTT 306
QY 349 LysSerGluValGlnGlnProValHisProLysProLysSerProAspSerArgAlaSer 368
Db 307 AAGAGTGAAGTTCACAGAGCTGTCCATCCCAAGCCCAAGTCCAGATTCACAGAGCTCC 366
QY 369 SerLeuSerGluSerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388

Db 367 AGTCTTTCTGAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 426
Qy 389 ThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGln 408
Db 427 ACCTCGGTGGAAATGTCAGAACACAGTCTATCCATGAGAGGCTCTCTTGGCCCAACACAG 486
Qy 409 ValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThr 428
Db 487 GTGTTTACATCAGCTGCTTCCTGCTCTTATGCAACACAACTCAGTCTAGGAACA 546
Qy 429 TyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448
Db 547 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCACCTCAATCAACTCTTTAAATCT 606
Qy 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLys 468
Db 607 AAGGCAACTATGATGAAGGCTTTGGCCACAGACACACAGGATCTATGGCAAGCAAA 666
Qy 469 AsnGluAsnGluGluLeuLeuArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 667 AATGAAACAGAGAGATTTGGAGAGACACCCAGCTTGCAATGCAAGGGAGACCCCT 726
Qy 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMet 508
Db 727 CACAGCAAGGGGTAGAAGATGCCCTATTGCTAAGGTGGTGTCTCTGGCTGCAAGTATG 786
Qy 509 GluAlaLysAlaSerSerGlnGlnGlnLysGluAspLysProAlaGluThrLysLysLeu 528
Db 787 GAAGCCCAAGCCCTCTCTCAGCAAGAGAGAGCAACAGCCAGCTGAACCAAGAGAGCTG 846
Qy 529 ArgIleAlaTyrProProThrGluLeuGlySerSerGlySerAlaLeuGluGly 548
Db 847 AGATGCCCTGGCCACCCCTCAGCACTTGAAGTTTCAGGAAGTGCCTTTGGAGGAGGG 906
Qy 549 IleLysMetSerLysProLysTyrProProGluAspGluLysSerLysProGluValPro 568
Db 907 ATCAAAATGTCAAGGCCAATGGCTCTCAGAGACGAATCAGCAAGCCCGAAGTTCTCT 966
Qy 569 GluAspValAspLeuAspLeuLysLysLeuArgSerSerSerLeuLysGluArgSer 588
Db 967 GAGGATCTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTCACTGAAGGAAAGAAC 1026
Qy 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 1027 CGCCCATCTACTGTAGCAGCTTCAATTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 1086
Qy 609 SerProPheLysArgLysGlyTyrProMetSerGluGlnSerGluSerValGlyGly 628
Db 1087 TCCCCACCTATCAGGAAGGCTGGAGCATGTCAGACGAGTGAAGAGTCTGTGGGTGA 1146
Qy 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnVal 648
Db 1147 AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGCTTCTAAGAGAAATGGAAATGTG 1206
Qy 649 GlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlu 668
Db 1207 GGAAGAAACAACTGGCAAAACAAAGAAATCTAAAGGAGAGACAGGGAAGAGATGAAGAA 1266
Qy 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 1267 GGTCAATAGTTGGAGATGAGAAATGAGATCTTGTAGAAAATGGTGACACTCCGATGAA 1326
Qy 689 AspAsnSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSer 708
Db 1327 GATGATAACAGCTTCTCAACACACATCTCCACAGAACCCCAAGTCTCTGAAATGGTGG 1386
Qy 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAsp 728
Db 1387 AGTTTTGTAGACAACACCTTTGCTGAAGAAATTCCTACTACTCAGAAATCCAGATCCAGAT 1446
Qy 729 ValGluLeuTrpGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArg 748
Db 1447 GTGGAACTCTGGAGGAGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAGAGA 1506

Qy 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 1507 AATCGTATTATGATGAGGATGAGGATGAAGAG 1539

RESULT 14

AAC59489
ID AAC59489 standard; cDNA; 2267 BP.
XX AAC59489;
XX
XX
DT 26-JAN-2001 (first entry)
XX
DE Human secreted protein gene 41 SEQ ID NO:51.

XX Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
KW antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulnary; gene therapy; autoimmune disease;
KW hyperproliferative disorder; neoplasm; cancer; cardiovascular disorder;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW infection; ocular disorder; wound healing; skin aging; food additive;
KW preservative; ss.
XX
XX Homo sapiens.

OS Homo sapiens.

XX WO200056755-A1.

XX 28-SEP-2000.

XX 16-MAR-2000; 2000WO-US06830.

XX 19-MAR-1999; 93US-0125361.

XX 10-DEC-1999; 93US-0169910.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-587661/55.

XX P-PSDB; AAB34132.

XX New isolated nucleic acid molecules encoding 49 human secreted proteins
PT used for preventing, treating or ameliorating medical conditions, for
PT diagnosing pathological conditions or as food additives or
PT preservatives

PS Claim 1; Page 357-358; 419pp; English.

XX The polynucleotide sequences given in AAC59449 to AAC59497 encode the
CC human secreted proteins given in AAB34092 to AAB34140. AAB34141 to
CC AAB34216 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissue
CC and cells the genes are expressed in. Examples of activities include:
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
CC cytostatic; cardiac; vasotropic; cerebroprotective; neurotropic;
CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
CC and vulnary. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases, hyperproliferative disorders e.g. neoplasms or
CC cancer of the breast or liver, cardiovascular disorders, cerebrovascular
CC disorders, angiogenesis, nervous system disorders, infections caused by
CC bacteria, viruses and fungi and ocular disorders. The polypeptides can
CC also be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used

CC in the exemplification of the present invention.
XX
SQ Sequence 2267 BP; 741 A; 416 C; 490 G; 617 T; 3 other;

Alignment Scores:
Pred. No.: 1586-104 Length: 2267
Score: 1586.00 Matches: 306
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.39% Indels: 0
DB: 21 Gaps: 0

US-09-890-549-4 (1-759) x AAC59489 (1-2267)

QY 454 GluGlyPheGlyHisArgProHisLysAspLeuTTPAlaSerLysAsnGluAsnGluGlu 473
DB 32 GAAGGCTTTGGCAGACACACAGGATCATGGCAAGCAAAATGAAACGAAGAG 91
QY 474 IleLeuGluAspProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyVal 493
DB 92 ATTTTGGAGACACAGCCAGCTTCAATGCAAGGGAGACCCCTCACAGCCCAAGGGTA 151
QY 494 GluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSer 513
DB 152 GAAGATGCCCTATTGCTAAGGTGGTGTCTCTGGCTGCAAGTATGGAAGCCCAAGCCCTCC 211
QY 514 SerGlnGlnGluGlySGLuAspLysProIleGluThrLysLysLeuArgIleAlaTTPPro 533
DB 212 TCTCAGCAGGAGAGAAAGACAGCAGCTGAACCAAGAGCTGAGGATCGCCCTGGCCA 271
QY 534 ProProThrGluLeuGlySerSerGlySerAlaLeuGluGlyLysMetSerLys 553
DB 272 CCCCCCATCTGAACCTTGGAGATTTCAGAGAGTCCCTTGGAGGAGGATCAAAATGTCAAG 331
QY 554 ProLysTTPProProGluAspGluLysSerLysProGluValProGluAspValAspLeu 573
DB 332 CCCAATGGCTCTGTAGAGCAAGATCAGACGCCAGCTTCTGAGATGTGATCTA 391
QY 574 AspLeuLysLysLeuArgArgSerSerLysLysGluArgSerArgProPheThrVal 593
DB 392 GATCTGAAGAGAGCTAAGACGATCTTCTCACTGAAGGAAAGAGCCGCCCATTCCTCTGA 451
QY 594 AlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProIleArg 613
DB 452 GCAGCTTCATTCAAGCACCCTGTCAAGAGCCCAAAATCTGTCTCCACCTATCAGG 511
QY 614 LysGlyTTPSerMetSerGluGlnSerGluGlnSerValGlyArgValAlaGluArg 633
DB 512 AAAGCTGGAGCATGTCAGACAGAGTGAAGAGTCTGTGGTGGAGAGATGTCAGAAAGG 571
QY 634 LysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTTP 653
DB 572 AAACAAGTGGAAATGCCAAGGCTTCTAAGAGAGATGGAGATGTGGGAAACAAACCTGG 631
QY 654 GlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGlu 673
DB 632 CAAACAAGAAATCTAAAGGAGACACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 691
QY 674 MetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPhe 693
DB 692 ATGAGAGATGAGATCTTTGAGAAATGGTGACAGACTCCGATGAAGATGAATACAGCTTC 751
QY 694 LeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTTPSerSerPheValAspAsn 713
DB 752 CTCACAAACAAATCTCCACAGAACCCCAAGTCTCTGAATTTGGTGGAGTTTGTAGACAAC 811
QY 714 ThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTTPGlu 733
DB 812 ACCTTTCTGAAGAATTCCTACTCAAGATCCAGAAATCCCAAGATGTGGAACTCTGGAG 871
QY 734 GlyGluValValLysGluLeuSerValGluGluGlnLysArgAsnArgTyrTyrAsp 753
DB 872 GGAGAGTGTCAAGAGCTCTCTGTGGAGAGACAGATAAGAGAAATCGTATTATGAT 931

QY 754 GluAspGluAspGluGlu 759
DB 932 GAGGATGAGGATGAGAG 949

RESULT 15
AAC93483
ID AAC93483 standard; cDNA; 2158 BP.
XX
AC AAC93483;
XX

DT 16-FEB-2001 (first entry)
XX
XX Human secreted protein gene 5 SEQ ID NO:15.

Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
neurotropic; neuroprotective; antibacterial; virucide; fungicide;
ophthalmological; vulnery; autoimmune disease; rheumatoid arthritis;
hyperproliferative disorders; cancer; cardiovascular disorder;
cardiac arrest; cerebrovascular disorder; nervous system disorder;
Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.

Homo sapiens.

WO2000061626-A1.

19-OCT-2000.

06-APR-2000; 2000WO-US09066.
09-APR-1999; 99US-0128698.
20-JAN-2000; 2000US-0176926.

(HUMA-) HUMAN GENOME SCI INC.
(ROSE/) ROSEN C A.
Rosen CA, Ruben SM, Komatsoulis G;
WPI; 2000-619227/59.
P-PSDB; AAB51831.

New nucleic acid molecules encoding 49 human secreted proteins for
diagnosing, preventing or ameliorating medical conditions and used for
food additives or preservatives -
Claim 1; Page 435; 516pp; English.

Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding
human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -
AAB51927 represent alternative polypeptides encoded by the genes, and
amino acid sequences with which they share homology. The genes and
proteins have activities dependent on the tissues and cells in which they
are expressed. Examples of their activities include immunosuppressive;
antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
virucide; fungicide; ophthalmological; and vulnery. The secreted
proteins, polynucleotides, antagonists and agonists may be useful in
treating, preventing and/or diagnosing diseases and disorders such as
autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
angioneurosis, nervous system disorders e.g. Alzheimer's disease,
infections caused by bacteria, viruses and fungi and ocular disorders
e.g. corneal infection. The polypeptides can also be used to aid wound
healing and epithelial cell proliferation, to prevent skin aging due to
sunburn, to maintain organs before transplantation, for supporting cell
culture of primary tissues, to regenerate tissues and in chemotaxis. The
polypeptides can also be used as a food additive or preservative to
increase or decrease storage capabilities, fat content, lipid, protein,
carbohydrate, vitamins, minerals, cofactors and other nutritional
components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are
used in the isolation and characterisation of the proteins and

CC polynucleotides of the invention.

XX Sequence 2158 BP; 714 A; 383 C; 458 G; 603 T; 0 other;

Alignment Scores:

Pred. No.: 5,62e-92 Length: 2158
Score: 1411.00 Matches: 286
Percent Similarity: 98.29% Conservative: 1
Best Local Similarity: 97.95% Mismatches: 0
Query Match: 35.93% Indels: 5
DB: 21 Gaps: 0

US-09-890-549-4 (1-759) x AAC93483 (1-2158)

```
QY 473 GluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGly 492
DB 1 GAGATTTTGAGAGAGACAGCCCACTTGCAAATGCAAGGGAGACCCCTCACAGCCAGGG 60
QY 493 ValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAla 512
DB 61 GTAGAAGATGCCCTATTGCTAAGGTGGTGTCTCGCTGCAAGTATGGAAGCCAGGCC 120
QY 513 SerSerGlnGlnLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTrp 532
DB 121 TCCTCTCAGCAGAGAGAGGAAGACAAGCCAGCTGAAACCAAGAAAGCTGAGGATCGCCTGG 180
QY 533 ProProThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyIleLysMetSer 552
DB 181 CCACCCCTCACTGAATTTGGAAGTTGGAAGTGCCTTGGAGGAAGGGATCAAAATGTCA 240
QY 553 LysProLysTrpProGluAspGluIleSerLysProGluValProGluAspValAsp 572
DB 241 AAGCCCAATGGCTCTCTGAGAGCAATCAGCAAGCCCGAAGTTCTGAGGATGTGGAT 300
QY 573 LeuAspLeuLysLysLeuArgSerSerSerLeuLysGluArgSerArgProPheThr 592
DB 301 CTAGATCTGAAGAAGTAAGACGATCTTCTCACTGAAGGAAAGAGCGCCCATTCACT 360
QY 593 ValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPhe 612
DB 361 GTAGCAGCTTCAATTTCAAGACACCTCTGTCAAGAGCCCAAAACCTGTGCCACCTATC 420
QY 613 ArgLysGlyTrpSerMetSerGluGlnSerGluSerValGlyGlyArgValAlaGlu 632
DB 421 AGGAAGGCTGGAGCATGTCAAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGAGTTGCAGAA 480
QY 633 ArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThr 652
DB 481 AGGAACAAGTGGAAATGCCAAGCTTCTAAGAGAAATGGGAATCTGGGAAACAACACC 540
QY 653 TrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeu 672
DB 541 TGGCAAAACAAGAATCTAAAGGAGAGACAGGGAAGAGTAAGGAAGGTCAAGTTTG 600
QY 673 GluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSer 692
DB 601 GAGATGGAGAAATAGAAATCTTGAGAAATGGTGAGACTCCGATGAAGATCATACAGC 660
QY 693 PheLeuLysGlnGlnSerPro--GlnGluProLysSerLeuAsnTrpSerSerPheVal- 711
DB 661 TTCTTCAACAACAATCTCCACCAAGAGAACCAAGTCTCTGAATTTGGTCGAGTTTGTGA 720
QY 712 AspAsnThrPhe-AlaGluGlu-PheThrThrGlnAsnGlnLysSerGlnAspValGluL 731
DB 721 GACAACACCTTTGGCTGAAGAATTTCACTACTCAGAATCAGAAATCCAGGATGTGAAC 780
QY 731 euTrpGluGlyValValLysGluLeuSerValGluGluIleLysArgAsnArgT 751
DB 781 TCTGGAGGGAGAGTGTGTCAAAGAGCTCTCTGTGGAAGACAGATAAGAGAAATCGGT 840
QY 751 yTrAspGluAspGluAspGluGlu 759
DB 841 ATTATGATGAGGATGAGGATGAGAG 866
```

Search completed: January 6, 2004, 22:38:06
Job time : 537 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 6, 2004, 22:07:02 ; Search time 3923 Seconds
(without alignments)
4702.293 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSPFNRQRTSLSLRVT.....LSVEQIKENRYDEDEEE 759

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US09890549/runat_06012004_094752_19831/app_query.fasta_1.903
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPECI=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=200 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09890549@cgn_1_2810/runat_06012004_094752_19831 -NCPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estov:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_eston:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_vrt:*
22: em_gss_vrt:*
23: em_gss_vrt:*
24: em_gss_vrt:*
25: em_gss_vrt:*
26: em_gss_vrt:*
27: em_gss_vrt:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2955	75.2	3242	11	AK049350 Mus muscu
2	2718	69.2	3108	11	AK031698 Mus muscu
3	2713	69.1	3100	11	AK028186 Mus muscu
4	2707	68.9	3160	11	AK085065 Mus muscu
5	1532.5	39.0	1075	12	BM909158 AGENCOURT
6	1401	35.7	1142	12	EM914155
7	1294	33.0	971	10	BE410108
8	1252.5	31.9	989	12	BM552304
9	1243	31.7	768	10	BG476496
10	1222	31.1	918	13	BU850203
11	1219.5	31.1	787	13	BU861039
12	1212	30.9	847	13	BU803212
13	1209	30.8	828	12	BI457843
14	1199.5	30.5	770	12	BI089829
15	1193	30.4	931	12	BM449793
16	1190	30.3	1120	10	BG120125
17	1186.5	30.2	993	12	BM458935
18	1186	30.2	884	14	CD171644
19	1181.5	30.1	951	10	BF528886
20	1172	29.8	715	9	AL048161
21	1170	29.8	765	12	BI557269
22	1148	29.2	941	10	BG112799
23	1126	28.7	643	13	BQ632628
24	1120.5	28.5	936	13	BQ713158
25	1114	28.4	884	13	BU527038
26	1106	28.2	634	10	BE729955
27	1103	28.1	777	28	AQ314576
28	1101	28.0	925	13	BU856577
29	1069	27.2	637	14	CA328969
30	1050.5	26.8	752	10	BE897604
31	1043	26.6	1095	10	BG112232
32	998	25.4	717	10	BF752641
33	991	25.2	765	14	CD512151
34	975	24.8	846	12	BG911181
35	969	24.7	747	14	CB963617
36	967.5	24.6	767	14	CB988778
37	959.5	24.4	931	13	BQ212132
38	955	24.3	728	13	BU631767
39	952.5	24.3	809	10	BE621212
40	945	24.1	841	10	BE542098
41	939	23.9	576	10	BE005092
42	935	23.8	933	10	BG539577
43	932	23.7	843	9	AA909018
44	906	23.1	665	12	BI153527
45	897.5	22.9	654	14	CB050200

ALIGNMENTS

RESULT 1

AK049350

LOCUS

DEFINITION

AK049350

VERSION

KEYWORDS

SOURCE

ORGANISM

Accession

Version

Keywords

Source

Organism

AK049350
Mus musculus ES cells cDNA, RIKEN full-length enriched library, full
clone: C330026L07 product: epithelial protein lost in neoplasm, full
insert sequence.

AK049350
Mus musculus
HTC, CAP trapper.
Mus musculus (house mouse)

AK049350
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

3242 bp mRNA linear HTC 05-DEC-2002

REFERENCE	1	Carninci, P. and Hayashizaki, Y.	
AUTHORS			
TITLE		High-efficiency full-length cDNA cloning	
JOURNAL		Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE		99279253	
PUBMED		10349636	
REFERENCE	2		
AUTHORS		Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL		Genome Res. 10 (10), 1617-1630 (2000)	
MEDLINE		20499374	
PUBMED		11042159	
REFERENCE	3		
AUTHORS		Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishimi, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
TITLE		RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	
JOURNAL		Genome Res. 10 (11), 1757-1771 (2000)	
MEDLINE		20530913	
PUBMED		11076861	
REFERENCE	4		
AUTHORS		Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Kachiwa, H., Quackenbush, J., Schiraldi, L. M., Staehli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, L., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S. and Hayashizaki, Y.	
TITLE		Functional annotation of a full-length mouse cDNA collection	
JOURNAL		Nature 409 (6821), 685-690 (2001)	
MEDLINE		21085660	
PUBMED		11217851	
REFERENCE	5		
AUTHORS		The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
TITLE		Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
JOURNAL		Nature 420, 563-573 (2002)	
REFERENCE	6		
AUTHORS		Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	
TITLE		Direct Submission	
JOURNAL		Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome	

COMMENT

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/.

FEATURES

Source

1. 3242
/organism="Mus musculus"
/mol_type="rRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C330026L07"
/db_xref="taxon:10090"
/clone="C330026L07"
/cell_type="ES cells"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
105. 2366
/note="unnamed protein product; epithelial protein lost in neoplasm (MGD|MG1:1920992, GB|NM_023063, evidence: BLASTN, 99%, match=3086)
putative"
/codon_start=1
/protein_id="BAC33699.1"
/db_xref="GI:26340072"
/translation="MESTFNRNROWTSLRVTAKELSLVKNKSSAIVEIFSKYQA
AEEANMERKQNPESLPQHFRRGLSVLKKWNPVAGAEFTDLPNSSEGGHTAD
YPAEAVTDKPAQVADREHTQKPRFGSRPEAVIOSRYPRSENSHDFKAQATESQK
MENCGLSRHEAEKPTGENTETSGKIEKYNVPLNRLKMPKEGHEQTSLNTQSRN
AVSQSSPASTYNELKTESKTHKEWKEVPPQPEAYSVHQEGSKYSTTSDMAKYQA
VPAEDTCNSQVSEAAQPMHPKPLSPDARTSLSPSSPSKTKAKFOAPAKESVCEQ
KVTYMERLANQVFIHISCFRCYCNKLSLGTAYSLHGRIVYCKPHNQLFKSKGNV
DEGFHGHKQDLWASKEDEETLGRPAQPPNAGESPHSPGVEDAPIAKVGLAASMEA
KASSORREDEKPAETKRLIAMPPEALGGSGSALEGIKVPKPKPEDDVCKTEAP
EDVDLDLKKLRSSSLKERSRPFTVAASFRITSSKPSASSPSLRKMGWSESESEEG
GGIATMERKQTNARPSKGNQWQGEVPRSDRSFELESENFMENGANIAE
DDNHVHAQSPGLEPFAEPAGSGFVDTTAAKEFTTQXQSDQVGFGEWVEVSELEQI
KRRRYDEDEDEE"

CDS

BASE COUNT 961 a 797 c 863 g 621 t
ORIGIN
Alignment Scores:
Pred. No.: 1.62e-218 Length: 3242
Score: 2955.00 Matches: 575
Percent Similarity: 83.55% Conservative: 60
Best Local Similarity: 75.66% Mismatches: 117
Query Match: 75.25% Indels: 8
DB: 11 Gaps: 3

US-09-890-549-4 (1-759) x AK049350 (1-3242)

Oy 1 MetCluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThra 20
Db 105 ATGGAATCGATCCATTTAATAGACGCCAGTCGACCTCCCTGTCATTGAGAGTAACAGCC 164
Oy 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 165 AAAGAGCTTTCTCTTGTCAACAAGAACAGTCATCCGCAATTGTGGAATAATTTCTCAAG 224
Oy 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 225 TACCAGAGGCTCTCTGAGAGGCCACATCGAAAGAGAGAAATAACCTGAAAGCTTG 284
Oy 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 285 CCCAGCACTTTAGAGAGGGACCCCTGTCTGTGTTAAAGAGAGTGGGAGAACCCGGT 344

QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluLeuArgHisArg 100
DB 345 GCTGGGGCAGATTCCACACAGACTCCTGCCAAACAGCAGCAGTGGGGTGGGCACACA 404
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
DB 405 GCGGACTACCTCCTGCTGAAGTACGGACAAAGCCTGCTCTCTGGAGTCAGACCTGACCGG 464
QY 121 GluGluGlnHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
DB 465 GAAGAGCACACCCAGCCCAACCTAGATTGGATCTCGTCCGGAAGCAGTATATCCAGAGC 524
QY 141 ArgTyrProHisLeuLeuGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
DB 525 CGGTATCTCGCTCAGAGAACACACAGATTTTAAAGCCCGCAGCAGGAGCCCAAAA 584
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluLeuSerGluAsn 180
DB 585 ATGGAATACTGTCTGGGATTCAGGCATGAAGCAGAGAACCCAGAGACGGCGCAAAAC 644
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
DB 645 ACAGAAACTTCAGGCAAAATAGAGAAATACAACTGCTTCCACTGAATAGACTGAAGATGATG 704
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
DB 705 TTGTGAAGAAGGTGAACACAAACCAAGAGTCTCTGGACCCAAAGCCGGAATGCGGGT 764
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluLeuGlyProGln 240
DB 765 GGAAGAGGCTCTGTGAATACTGTTCCCTGGATGACTGGGAATAGGTGCGGGTCAAT 824
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
DB 825 TTGTCTCTCTGCTCACTCACTCGGAGAAACAGAGAGTAAAGAGAACTCTCGAGCTGCCA 884
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaValSerLys 280
DB 885 CCCCTCTCAGAACTCTCATAAAGACCGCATGGCCAAAGTACCAGCTGTCAGTGTCAAAG 944
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLysAlaSerGlyGlyGluLysIle 300
DB 945 CAGAGCACCCAGCAGCTACACAATAGCTGTAACCACTGTTT-----GAAAGCAAACT 998
QY 301 HisLysMetGluGlnLysGluAsnValProGlyProGluValCysIleThrHisGln 320
DB 999 CATAAATGGGAACAGAGAGAAATGTGCCCCAGGTCCCGAGGCTACAGGCTCCATCAG 1058
QY 321 GluGlyGluLysIleSerAlaAsnGluSerLeuAlaValArgSerThrProAlaGlu 340
DB 1059 GAAGGAAGCAAGGTTCTTACAACTGAGAATAGTCTAGTGGCCCTCTCTGTCCTGCTGAA 1118
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
DB 1119 GATGACACCTGTAATCTCCAGGTGAAGAGCAGGCCAGCAGCCTATGCAACCTTAAGCCG 1178
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
DB 1179 CTGAGTCTCGAGCCAGGACCTCCAGCTCTCCGGAAGTCTCTCTCCAAACAGCGAAG 1238
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
DB 1239 AAGTTTCAGGCGCGGCAAGAGAGAGTGGTGGAGTGTGAGAAGAGCGGTGACCCCATG 1298
QY 401 GluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
DB 1299 GAACGGCTCTCGGCAACCCAGCAGGTGTTTACATCAGCTGTTTCCGATCTCTACTGC 1358
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
DB 1359 AACACAGCTCAGCTAGAGAACATATGCATCTTATATGAGCAATCTACTGTAAGCCT 1418

QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
DB 1419 CACTTCAATCAACTCTTCAAGTCCAAAGGCAACTATACGAGGGCTTTGGGCACAAGCAG 1478
QY 461 HisLysAspLeuTyrAlaSerLysAsnGluAsnGluGluLeuGluArgProAlaGln 480
DB 1479 CATAGAGATCTGTGGCAAGCAAGAGCGACAATGAGGAGACTTTTGGGAGACCGAGCCAG 1538
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
DB 1539 CCGCCTAATGCAGGGAGAGCCCCCATAGCCGGGTGTAGAAGATGCCCCATCGCCCAAG 1598
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
DB 1599 GTGCGCGTGTGGCGCAAGTATGGAAGCAAGCCCTCTTCTCAGAGGGAGAGGAGAT 1658
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTyrProProProThrGluLeuGlySer 540
DB 1659 AAGCCTCGGAGACCAAGAGCTGAGGATCGCTGGCTCCGCCAGCGAGCTGGCGGT 1718
QY 541 SerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTyrProProGluAsp 560
DB 1719 TCCGGAAGTCCCTGGAGGAGGAGATCAAGTATCGAAGCCCAAGTGGCTCCGAGGAT 1778
QY 561 GluLeuSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
DB 1779 GAGCTCTGCAAGCGGAGGCCCCCGGAGGATGTAGATCTGACCTGAAGAAGCTACGGCG 1838
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
DB 1839 TCTTCGTCCTGAAAGAGAGGAGCCCGCTTTACGTTAGCAGCTTCTTTCGAACCTCT 1898
QY 601 SerValLysSerProLysThrValSerProProLysArgLysGlyTyrSerMetSerGlu 620
DB 1899 TCCATCAAGAGCCCAAGGCTCTTCCCATCTCTCAGGAAAGGTTGGAGCAGTCCGAG 1958
QY 621 GlnSerGluGluSerValGlyGly---ArgValAlaGluArgLysGlnValGluAsnAla 639
DB 1959 CAGAGTGAGGAGTTTGGAGAGGAGTAGCGACGATGGAAGGAAACAAACGGAAATGCC 2018
QY 640 LysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTyrGlnAsnLysGluSerLys 659
DB 2019 AGACCTCTGGGAGAGAGAGAAATGTGGAAATACCGCTGGCAG----- 2063
QY 660 GlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluLeu 679
DB 2064 GCGAGAGAGTGGCGAGGAGCAAGGACCGTAGTCTCTTTGAGCTGAGAGTGAAGATTTT 2123
QY 680 ValGluAsnGlyAlaAspSerSerAspGluAspAspAsnSerPheLeuLysGlnSerPro 699
DB 2124 ATGGAATAATGGAGCAACATAGCTGAAGATGACACCATGTCCACGACGAGTCCCGG 2183
QY 700 GlnGluProLysSerLeuAsnTyrSerSerPheValAspAsnThrPheAlaGluGluPhe 719
DB 2184 TTAGAACCAGGCTCCGGGTGGTCTGGCTTTGTAGACACCGCTGCTTAAGAATTC 2243
QY 720 ThrThrGlnAsnGlnLysSerGlnAspValGluLeuTyrGluGlyValValLysGlu 739
DB 2244 ACTACCCAGATCAAAATCCCAAGATGCGGTTCTGGAGGGAGAGTGGTCCGAGAG 2303
QY 740 LeuSerValGluGluGlnLysLysArgAsnArgTyrTyrAspGluAspGluGlu 759
DB 2304 CTGTCTGTGGAGAACAGATAAAGAGAAACCGGTACTATGATGAGGACGAGGATGAAGAA 2363

RESULT 2

AK031698

LOCUS

DEFINITION

Mus musculus 13 days embryo male testis cDNA, RIKEN full-length

enriched library, clone:6030492D07 product:epithelial protein lost

in neoplasm, full insert sequence.

ACCESSION

AK031698

VERSION

AK031698.1

GI:26327552

KEYWORDS

HTC; CAP trapper.

AK031698 3108 bp mRNA linear

Mus musculus 13 days embryo male testis cDNA, RIKEN full-length

enriched library, clone:6030492D07 product:epithelial protein lost

in neoplasm, full insert sequence.

ACCESSION

AK031698

VERSION

AK031698.1

GI:26327552

KEYWORDS

HTC; CAP trapper.

QY 116 AlaLysAlaAspGlnGluGlnHisProArgSerArgLeuArgSerProGlu 135
Db 316 GTGAGAGTACCGGGAAGACACACCCAGCCCAACCTAGATTGATCTGTCGGAA 375
QY 136 AlaLeuValGlnGlyArgTyrProHisIleLysAspGlyGluAspLeuLysHisSer 155
Db 376 GCAGTTATCCAGAGCCGGTATCTCGCTCAGAGACAGACCCAGCAATTTAAAGCCAGGCC 435
QY 156 ThrGluSerLysLysMetGluAsnGlnLysGluSerArgHisGluValGluLysSer 175
Db 436 ACCGAGAGCCAAAATGGAAAATCTGTCTGGGAGATTCCAGGCATGAAGCAGAGGCCA 495
QY 176 GluIleSerGluAsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsn 195
Db 496 GAGACGAGCGAAACACAGAACTTCAGGCATAATAGAGAAATACAACTTCACCTGAAT 555
QY 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGln 215
Db 556 AGACTGAAGATGATGTTTGGAGAAAGGTGAACCAACCAACCAAGAGTCTCTGGACCCAA 615
QY 216 SerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGlu 235
Db 616 AGCCGAAATCGCGGTGGAAGGAGCTCTCTGAAAACAACTGTTCCTGTGATGACCTGGAA 675
QY 236 IleGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGlnSerArg 255
Db 676 ATAGTGGCGGTTCATTGTCTCTCTGCACTCACTCGAGAAACACAGAGTAGAGG 735
QY 256 AsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGln 275
Db 736 AATCTGGAGTGCACCGCTCTCAGAAACCTCCATAAAGGACCGCATGGCCAGTACCAG 795
QY 276 AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly 295
Db 796 GTTGCACTTCCAGCAGAGCAGCCAGCCAGCTACAAATGAGCTGAAAACCACT-- 852
QY 296 GlyGluIleLysIleHisLysMetGluGlnLysGluAsnValProGlyProGluVal 315
Db 853 ---GAAAGCAAACTCATTAATGGGAACAGAGGAGAAATGTGCCCCAGGTCCCGAGGCC 909
QY 316 CysIleThrHisGlnGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArg 335
Db 910 TGCAGCGTCCATCAGGAAGAGCAAGGTTCCTCAACTGAGATGATCTAGTGGCCCTC 969
QY 336 SerThrProAlaGluAspSerArgAspSerGlnValLysSerGluValGlnGlnPro 355
Db 970 TCTGTCCCTGCTGAAGATGACACCTGTAACCTCCAGGTGAAGAGGAGGCCAGCAGCT 1029
QY 356 ValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerPro 375
Db 1030 ATGCACCTTAAGCCGCTGAGTCTCCAGCCAGGACCTCCAGTCTTCCGGAAGTCTCT 1089
QY 376 ProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLys 395
Db 1090 TCCAAAACAGCGAAGATTTTCAGCGCGCGGCAAGAGGAGAGCTCGGTGGAGTGCAGAG 1149
QY 396 ThrValTyrProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPhe 415
Db 1150 ACGGTGTACCCCATGAAACGGCTCTGCGCAACAGCAGAGGTGTTTCATCATCAGTGTTC 1209
QY 416 ArgCysSerTyrCysAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArg 435
Db 1210 CGATGCTCCTACTGCAACCAAGCTCAGTCTAGGAACATATGCATCTCTACATGGACGA 1269
QY 436 IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGly 455
Db 1270 ATCTACTGTAGCCCTCACTTCATCACTCTTCAAGTCCAAAGGCAACTATGACGAGGC 1329
QY 456 PheGlyHisArgProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeu 475
Db 1330 TTGGGCACAGCAGCATTAAGATCTGTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTTG 1389
QY 476 GluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAsp 495

Db 1390 GGGAGACAGCCAGCCCTTAATGCGGGAGAGCCCTTAGCCGGGTGTAGAAGAT 1449
QY 496 AlaProfileAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGln 515
Db 1450 GCCCCCATCGCAAGTCTGGCGTCTGGCGCAAGTATGGAAGCCCAAGCCCTCTTCTCAG 1509
QY 516 GlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTrpProPro 535
Db 1510 AGGAGAGAGGAGGATAGCTCGGAGACCAAGAGCTGAGGATCGCTCGCTCCCGCCCA 1569
QY 536 ThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyLysMetSerLysLysProLys 555
Db 1570 GCCAGAGCTGGCGGTTCGGAAGTCCCTGAGGAAAGGATCAAGATATCAAGCCCAAG 1629
QY 556 TrpProGluAspGluIleSerLysProGluValProGluAspValAspLeuAspLeu 575
Db 1630 TGGCTCCGAGGATGAGCTTCGAGAGGAGGCCCGGAGGATGATAGTCTCGACCTG 1689
QY 576 LysLysLeuArgArgSerSerSerLysGluArgSerArgProPheThrValAlaAla 595
Db 1690 AAGAAGCTGCGCGGTCTTCGTCTCTGAGAGGAGGAGCCCGCTTACGTTAGCAGCT 1749
QY 596 SerPheGlnSerThrSerValLysSerProLysThrValSerProLysLysGly 615
Db 1750 TCCTTTCCAACTCTTCATCAAGAGCCCAAGCCCTCTTCGCCATCTCTCAGGAAAGGT 1809
QY 616 TrpSerMetSerGluGlnSerGluLysSerValGlyGly---ArgValAlaGluArgLys 634
Db 1810 TGGAGCAGTCCGAGCAGAGTGGAGGTTTGGAGGAGGATAGCGACGATCGAAGGAAA 1869
QY 635 GlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGln 654
Db 1870 CAACGGAATATGCCAGCCCTCTGGGAGAAAGAGAAATGTGGAAAATCAGCTGCGAG 1929
QY 655 AsnLysGluSerLysGlyThrGlyLysArgSerLysGluGlyHisSerLeuGluMet 674
Db 1930 -----GCCGAGGAAGTCCCGAGGAGCAGGAGCCAGTCTTCTTTGAGCTG 1974
QY 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeu 694
Db 1975 GAGAGTGAAGATTTATGGAATAATGGAGCAACATAGCTGAAGATGACCAACCATGTCAC 2034
QY 695 LysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThr 714
Db 2035 GCACAGCAGTCCCGGTAGAACCCGAGGCTCCGGGTGGTCTGGCTTTGTAGACACACG 2094
QY 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGluAspValGluLeuTrpGluGly 734
Db 2095 GCTGCTAAAGATTCATCCCAAGATCAAAATCCCAAGATGTCGGGTTCCTGGGAGGA 2154
QY 735 GluValValLysGluLeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGlu 754
Db 2155 GAAGTGTCTCGAGAGCTGTCTGTGGAGAACAGATTAAGAGAAACCGGTACTATGATGAG 2214
QY 755 AspGluAspGluGlu 759
Db 2215 GACGAGGATGAAGAA 2229
RESULT 3
AK028186
LOCUS
DEFINITION
Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
enriched library, clone:2610034J22 product:epithelial protein lost
in neoplasm, full insert sequence.
AK028186
VERSION
AK028186.1 GI:26389833
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1

Db 373 GCAGTTATCCAGAGCGGTATCTCGCTCAGAGAACAGCCACGATTTTAAAGCCAGGCC 432
Qy 156 ThrGluSerLysLysMetGluAsnCysLeuGlyCysLeuArgHisGluValGluLysSer 175
Db 433 ACGGAGGCCAAATAATGGAATACTGTCTGGAGATTCAGGCATGAAGCAGAGAACCA 492
Qy 176 GluIleSerGluAsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsn 195
Db 493 GAGACGAGCCAAACACAGAACTTCAGGCAAAATAGAGAAATACACGTTCCACTGAAT 552
Qy 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGln 215
Db 553 AGACTGAAGATGATGTTTGAGAAAGGTGAACACAAACCAAGAGTCTCTGGACCCAA 612
Qy 216 SerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGlu 235
Db 613 AGCCGAATTCGGGTGGAGAGGCTCTCTGAAAAACAATGTTCTCCCTGGATCAGTGGAA 672
Qy 236 IleGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgArg 255
Db 673 ATAGGTGCGGGTCATTGTTCTCTCTCGATTCAACTCGGAGAAAACAGAGTAAGAG 732
Qy 256 AsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGln 275
Db 733 AATCTGGAGTCGCCAGCCTCTCAGAAACCTCATAAAGGACCGCATGSCCAAGTACCAG 792
Qy 276 AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly 295
Db 793 GTCTGAGTGTCCAGCAGAGCAGCCAGCCAGCTACACAAATGAGCTGAAACCAAGT... 849
Qy 296 GlyGluIleLysIleHisLysMetGluGlnLysGluAsnValProGlyProGluVal 315
Db 850 ---GAAAGCAAACTCATAAATGGGACAGAGGAGAAATGTCCTCCAGTCCCCAGGCC 906
Qy 316 CysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluSerLeuAlaValArg 335
Db 907 TGCAGCGTCCATCAGAGCAAGCAAGGTTTCTACAACTGAGAAATAGTCTAGTGGCCCTC 966
Qy 336 SerThrProAlaGluAspSerArgAspSerGlnValLysSerGluValGlnGlnPro 355
Db 967 TGTGTCCTCTGCTGAAGATGACACCTGTAACTCCAGGTGAAGAGCAGGCCAGCAGCCT 1026
Qy 356 ValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerPro 375
Db 1027 ATGCACCTTAACCGCTAGTCTCTGACGCCAGGACCTCCAGTCTCCGGAAGTTCTCCT 1086
Qy 376 ProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLys 395
Db 1087 TCCAAACAGCGAAGAGTTTCAGGCGCGCGCAAGAGAGAGCTGCTGGATGTCAGAG 1146
Qy 396 ThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPhe 415
Db 1147 ACGGTGTACCCCATGGAACGGCTCTGCGCAACAGCAGAGGTGTTTTCATCAGCTGTTTC 1206
Qy 416 ArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArg 435
Db 1207 CGATCTCTCTACTGCAACAAACAGCTCAGTCTAGGAACATATGCATCTTACATGGACGA 1266
Qy 436 IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGly 455
Db 1267 ATCTACTGTAGCCTCACTTCACTCACTCTTCAAGTCCAAAGGCAACTATGACGAGGCG 1326
Qy 456 PheGlyHisArgProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeu 475
Db 1327 TTTGGGCAACAGCAGCATAAGATCTGTGGGCAAGCAGCAGCAATGAGGAGACTTTG 1386
Qy 476 GluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAsp 495
Db 1387 GGGAGACAGCCAGCCAGCCCTCAATGAGGGGAGAGCCCAATAGCCCGGTGTAGAGAT 1446
Qy 496 AlaProIleAlaLysValGlyValIleAlaAlaSerMetGluAlaLysAlaSerSerGln 515
Db 1447 GCCCCCATCGCAAGGTGCGGCTGTGCGGCAAGTATGGAAGCCAAAGGCTCTCTCTCAG 1506

Qy 516 GlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaIleTrpProPro 535
Db 1507 AGGGAGAGGAGGAGGATTAAGCTCGGAGACCAAGAAGCTGAGGATCGCTGCGCTCCGCCA 1566
Qy 536 ThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLys 555
Db 1567 GCCGAGCTGGGCGGTTCGGAAGTGCCTCGGAGAGGAGATCAAAGTATCGAAGCCCAAG 1626
Qy 556 TrpProProGluAspGluIleSerLysProGluValProGluAspValAspLeuAspLeu 575
Db 1627 TGGCTCCGAGGAGTACGCTCGCAAGCAGGAGGCCCGGAGGATGTAGATCTTGACCTG 1686
Qy 576 LysLysLeuArgArgSerSerSerLysLysGluArgSerArgProPheThrValAlaAla 595
Db 1687 AAGAAGCTGCGCGGTCTTCGTCCTGAAGGAGAGGAGCCCGCTTTACGTAGCAGCT 1746
Qy 596 SerPheGlnSerThrSerValLysSerProLysThrValSerProProIleArgLysGly 615
Db 1747 TCCCTTGAACCTCTTCATCAAGAGGCCCAAGGCCCTCTTCGCCATCTCTCAGAAAGGT 1806
Qy 616 TrpSerMetSerGluGlnSerGluGluSerValGlyGly---ArgValAlaGluArgLys 634
Db 1807 TGGAGCAGTCCGAGCAGAGTGAAGGTTTGGAGGAGGATAGCGACGATGGAAGGAAA 1866
Qy 635 GlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGln 654
Db 1867 CAACGGAAAAATGCCAGACCTCTCTGGGAGAAAGAGAAATGTGGAAAAATCACGCTGGCAG 1926
Qy 655 AsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlyHisSerLeuGluMet 674
Db 1927 -----GCCGAGGAAGTCCGAGGAGCAAGCAGCGTAGTCTCTTTTGAGCTG 1971
Qy 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeu 694
Db 1972 GAGAGTGAAGATTTATGGAATAATGGAGCAACATAGCTGAAGATGACAAACCATGTCCAC 2031
Qy 695 LysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThr 714
Db 2032 GCACAGCAGTCCCCGTAGAACCCGAGGCTCCGGTTGGTCTTGGCTTTGTAGACACACG 2091
Qy 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly 734
Db 2092 GCTGCTAAAGATTTACTACCCACATCAAAATCCCAAGATGTCGGGTTCTGGGAGGGA 2151
Qy 735 GluValValLysGluLeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGlu 754
Db 2152 GAAGTGTCTCGAGAGTGTCTGTGGAGAACACAGATAAGAGAAACCGGTACTATGATGAG 2211
Qy 755 AspGluAspGluGlu 759
Db 2212 GACGAGGATGAAGAA 2226
RESULT 4
AK085065 3160 bp mRNA linear HTC 05-DEC-2002
LOCUS Mus musculus 13 days embryo lung cDNA, RIKEN full-length enriched
DEFINITION library, clone:D430032I03 product:epithelial protein lost in
neoplasm, full insert sequence.
ACCESSION AK085065
VERSION AK085065.1 GI:26351432
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2

QY 176 GluIleSerGluAsnThrAspAlaSerGlyLysIleGluLysThrAsnValProLeuAsn 195
 DB 546 GAGACAGGCGAACAACAGAACTTCAGGCAAAATAGAGAAATACAACTTCACCTGAAT 605
 QY 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGln 215
 DB 606 AGACTCAAGATGATTTTCAGAAAGGTGAACACCAACCAAGAGTCTCTGGACCCAA 665
 QY 216 SerArgSerAlaSerGlyArgIleSerGluAsnSerTySerLeuAspAspLeuGlu 235
 DB 666 AGCGGAAATCGGGTGAAGAGGCTCTCTGAAAACAACTGTCTCCGTGATCAGTGGGAA 725
 QY 236 IleGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArg 255
 DB 726 ATAGGTGCGGTCAITTTGCTCTCTGCACTCAACTCGGAGAAACAGAGTAAAGAG 785
 QY 256 AsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyGln 275
 DB 786 AATCTGAGCTGCCAGCTCTCAGAACTCTCAATAAGAGCGCATGGCCAACTACAG 845
 QY 276 AlaAlaValSerLysGlnSerSerThrAsnTyThrAsnGluLeuLysAlaSerGly 295
 DB 846 GCTGCAGTGTCCAAAGCAGAGCAGCCAGCCAGCTACACAAATGAGTGAACCAAGT--- 902
 QY 296 GlyGluLysIleHisLysMetGluGlnLysGluAsnValProProGlyProGluVal 315
 DB 903 ---GAAAGCAAACTCATAAATGGGAACAGAGGAATGTGCCCCAGGTCCCGAGGCC 959
 QY 316 CysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArg 335
 DB 960 TCGAGCTCCATCAGGAAGAGCAAGAGTTCTCAACTGAGATAGTCTAGTGGCCCTC 1019
 QY 336 SerThrProAlaGluAspAspSerArgAspSerGlnValLysSerGluValGlnGlnPro 355
 DB 1020 TCTGCTCCCTGCTGAAGTACACCTCTTAACCTCCAGGTGAAGAGCGAGCCAGCAGCCT 1079
 QY 356 ValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerPro 375
 DB 1080 ATGCACCTTAAGCCGTAGTCTCTGACGCCAGCAGCCCGAGTCTCCGGAAAGTTCTCT 1139
 QY 376 ProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLys 395
 DB 1140 TCCAAACAGCGAAGAGTTTCAGCGCGCGGCAAGAGAGAGCTGCGTGGAGTGCAGAG 1199
 QY 396 ThrValTyProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPhe 415
 DB 1200 ACGGTGTACCCCATGGAAGCGCTCTGCCCACACAGAGGTGTTTCATCAGCTGTTTC 1259
 QY 416 ArgCysSerTyCysAsnAsnLysLeuSerLeuGlyThrTyThrAlaSerLeuHisGlyArg 435
 DB 1260 CGATGCTCTACTGCAACAACAAAGCTCAGTCTAGGAACATATGCATCTTACATGGACGA 1319
 QY 436 IleTyCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyAspGluGly 455
 DB 1320 ATCTACTGTAGCCCTCACTTCAATCACTCTTCAAGTCCAAAGGCAACTATCAGCAGAGC 1379
 QY 456 PheGlyHisArgProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluLysLeu 475
 DB 1380 TTGGGCACAGCACATAGAGATCTGGGCAAGCAAGAGCGCAACTAGGAGACATTTG 1439
 QY 476 GluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAsp 495
 DB 1440 GGGAGACAGCCAGCCGCTAATCGAGGAGAGCCCAATAGCCGGGTGTAGAGAT 1499
 QY 496 AlaProIleAlaLysValGlyValLeuAlaLysMetGluAlaLysAlaSerGln 515
 DB 1500 GCCCCATCGCAAGTGGGTGCTGGCGCAAGTATGGAAGCAAGGCCCTCTTCTCAG 1559
 QY 516 GlnGluLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTrpProPro 535
 DB 1560 AGGGAGAGGGAGGATAAGCCTCGGAGACCAAGAGCTGAGGATCGCTGCCCTCCGCCA 1619
 QY 536 ThrGluLeuGlySerSerGlySerAlaLeuGluGlyIleLysMetSerLysProLys 555

DB 1620 GCCGAGCTGGCGGTTCCGGAAGTCCCTCGAGGAAGGGATCAAAAGTATCGAAACCCAAAG 1679
 QY 556 TrpProGluAspGluIleSerLysProGluValProGluAspValAspLeuAspLeu 575
 DB 1680 TGGCTCCGAGGATGAGCTGCAACAGGAGGCCCGGAGGATGATGATCTCGACCTG 1739
 QY 576 LysLysLeuArgArgSerSerSerLeuLysGluLysArgSerArgProPheThrValAlaAla 595
 DB 1740 AAGAAGCTGGCGGCTCTTCGTCTCCCTGAAGAGGAGGAGCCCGCTTACGGTAGCAGCT 1799
 QY 596 SerPheGlnSerThrSerValLysSerProLysThrValSerProProLysLeuGly 615
 DB 1800 TCCCTTTGAACTCTTCATCAAGAGGCCCAAGGCCCTCTTCGCCATCTCTCAGGAAAGGT 1859
 QY 616 TrpSerMetSerGluGlnSerGluLysSerValGlyGly---ArgValAlaGluArgLys 634
 DB 1860 TGGAGCAGTCCGAGCAGAGTGGAGGTTTGGAGGAGGATAGCGACGATGGAAGGAAA 1919
 QY 635 GlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrTrpGln 654
 DB 1920 CAACGGGAAATGTCACACCTCTGGGAGAAAGAGAAATGTGGGAAATCAGCGTGCAG 1979
 QY 655 AsnLysGluSerLysGlyThrGlyLysArgSerLysGluGlyHisSerLeuGluMet 674
 DB 1980 -----GGCGAGGAAGTCCCGAGGAGCAAGGACCGTAGTCTTTTGACCTG 2024
 QY 675 GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeu 694
 DB 2025 GAGAGTGAAGATTTATGGAATGAGCAACATAGCTGAAGATGACCAACCATGTCCAC 2084
 QY 695 LysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThr 714
 DB 2085 GCACAGCAGTCCCGTTAGAACCCGAGGCTCCCGGTTGGTCTGCGCTTTGTAGACACAG 2144
 QY 715 PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly 734
 DB 2145 GCTGCTAAAGATTCATCTACCCAGATCAAAATCCCAAGATGTCGGGTTCTGGAGGGA 2204
 QY 735 GluValValLysGluLeuSerValGluGluGlnLysArgAsnArgTyTrpAspGlu 754
 DB 2205 GAAGTGTCTCGAGAGCTCTCTGTGGAGGAACAGATAAAGAGAACCCGTACTATGATGAG 2264
 QY 755 AspGluAspGluGlu 759
 DB 2265 GACGAGGATGAAGAA 2279
 RESULT 5
 BM909158
 LOCUS
 DEFINITION BM909158 1075 bp mRNA linear EST 12-MAR-2002
 5', mRNA sequence.
 ACCESSION BM909158
 VERSION BM909158.1 GI:19359537
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1075)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM1904 row: j column: 10

High quality sequence stop: 696.

```

FEATURES
  source
    1..1075
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:5431785"
      /tissue_type="astrocytoma grade IV, cell line"
      /lab_host="DH10B (phage-resistant)"
      /clone_lib="NIH_MGC_98"
      /notes="Organ: brain; Vector: pOTB7; Site: 1: XhoI; Site 2:
      EcoRI; cDNA made by oligo-dt priming. Directionally
      cloned into EcoRI/XhoI sites using the following 5'
      adaptor: GGCACGAG(G). Library constructed by Ling Hong
      in the laboratory of Gerald M. Rubin (University of
      California, Berkeley) using ZAP-cDNA synthesis kit
      (Stratagene) and Superscript II RT (Life Technologies).
      Note: this is a NIH MGC Library."
BASE COUNT      354 a   256 c   258 g   207 t
ORIGIN
Alignment Scores:
Pred. No.:      8,7e-109      Length:      1075
Score:          1532.50      Matches:     312
Percent Similarity: 95.78%      Conservative: 6
Best Local Similarity: 93.98%      Mismatches: 11
Query Match:     39.02%      Indels:      4
DB:              12          Gaps:         1

US-09-890-549-4 (1-759) x BM909158 (1-1075)
QY      80 GlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHis 99
Db      1 GGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAAACAGCAGCACTGAGATTAGGCAC 60
QY     100 ArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAsp 119
Db     61 AGAGCAGACCATCTCTCTGCTGAAGTGAACAAGCCAGCTGCTTCTGGAGCCAAAGCTGAC 120
QY     120 GlnGluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGln 139
Db     121 CAAGAAAGAAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAGCCCTCTGTTAG 180
QY     140 GluArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLys 159
Db     181 GGTTCGATATCCCCACATCAAGACCGGTGAGGATCTTAAAGACCACTCAACAGAAAGTAA 240
QY     160 LysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGlu 179
Db     241 AAAATGGAAAATTGTCTAGAGAAATCCAGGCAATGAAGTAGAAAAATCAGAAATCAGTGAA 300
QY     180 AsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMet 199
Db     301 AACACAGATGCTTCGGGCAAAATAGAGAAATATATATGTTCCGCTGAACAGAGCTTAAGATG 360
QY     200 MetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAla 219
Db     361 AAGTTTGAAAAGGTTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAGTGCA 420
QY     220 SerGlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGly 239
Db     421 ACTGAAGGAAGATCTCTGAAACACAGCTATTCTCTAGATGACCTGGAAATAGGCCAGGT 480
QY     240 GlnLeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeu 259
Db     481 CAGTTGTCTATCTTCTACATTTGACTCGGAGAAAATAGAGTAGACGAAAATCTCGAATCT 540
QY     260 ProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSer 279
Db     541 CCACGCCCTCTCAGAAACCTCTATAAAGATCGAATGGCCAAAGTACCAGGACGCTGTCTCC 600
QY     280 LysGlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluLys 299

```

```

Db      601 AAAAAGACGCTCAACCAACTATACAAATGAGCTGAAAGCGAGTGGTGGCGAAATCAAA 660
QY      300 IleHisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHis 319
Db      661 ATTCAATAATGGAGCAAAAGAGATGTGCCCCAGGTCTCTGAGGTCTGATCACCCT 720
QY      320 GlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAla 339
Db      721 CAGGAAGGGGAAAAGATTCTGCAATAGCAATAGCCTGGCAGTCCGCTTCCACCCCTGCC 780
QY      340 GluAspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLys 359
Db      781 GAAGATGACTCCCGTGACTCCAGGTTAAGAGTCAAGGTTCAACAGCCTGTCCATCCCAGG 840
QY      360 ProLeuSerProAspSerArg-AlaSerSerLeuSerGluSerSerProProLysAlaMe 379
Db      841 CCACCTAGTCAGATTCAGAAAGCTTCAGTCTTCTGAAAAGTCTCTCTCCCAAGA-AT 899
QY      379 tLysLysPheGlnAlaProAlaArgGluThrCys---ValGluCysGlnLysThrValTy 398
Db      900 GGAAGATTTCGGGACCTCGAGAAAGACCTGGGGGGAATGCGCAAAAACACACTCA 959
QY      398 rProMetGlu-ArgLeuLeuAlaAsnGlnGln 408
Db      960 TCCCATGGAAGCTCTCTTTGGCAACAAGAA 991

RESULT 6
BM914155
LOCUS      BM914155
DEFINITION      BM914155 1142 bp mRNA linear EST 12-MAR-2002
AGENCOURT 6608184 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5479040
5', mRNA sequence.
ACCESSION      BM914155
VERSION        BM914155.1 GI:19364534
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
REFERENCE      1 (bases 1 to 1142)
AUTHORS        NIH-MGC http://mgi.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-x@mail.nih.gov
                Tissue Procurement: ATCC
                cDNA Library Preparation: Rubin Laboratory
                DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
                Cloning by: Agencourt Bioscience Corporation
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LCM1999 row: k column: 09
                High quality sequence stop: 678.

FEATURES
  source
    1..1142
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:5479040"
      /tissue_type="astrocytoma grade IV, cell line"
      /lab_host="DH10B (phage-resistant)"
      /clone_lib="NIH_MGC_98"
      /notes="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
      EcoRI; cDNA made by oligo-dt priming. Directionally
      cloned into EcoRI/XhoI sites using the following 5'
      adaptor: GGCACGAG(G). Library constructed by Ling Hong
      in the laboratory of Gerald M. Rubin (University of
      California, Berkeley) using ZAP-cDNA synthesis kit
      (Stratagene) and Superscript II RT (Life Technologies).
      Note: this is a NIH MGC Library."
BASE COUNT      376 a   282 c   276 g   208 t
ORIGIN

```

Alignment Scores:
 Pred. No.: 1.49e-98 Length: 1142
 Score: 1401.00 Matches: 307
 Percent Similarity: 89.24% Conservative: 8
 Best Local Similarity: 86.97% Mismatches: 14
 Query Match: 35.68% Indels: 24
 DB: 12 Gaps: 4

US-09-890-549-4 (1-759) x BM914155 (1-1142)

QY 71 ValLeuLysLysLysTrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeu 90
 DB 1 GTGTTAAAGAAAGTGGGAGAACCCAGGGCTGGGAGAGAGTCTCACACAGACTCTCTA 60

QY 91 ArgAsnSerSerThrGluLeuArgHisArgAlaAspHisProAlaGluValThrSer 110
 DB 61 CGGAACAGCAGCAGTCTAGATTAGGCACAGAGCAGACCATCTCTCTGCTGGAAGTGAAGC 120

QY 111 HisAlaAlaSerGlyAlaLysAlaAspGlnGluGlnGluInHisProArgSerArgLeu 130
 DB 121 CACGCTGCTCTTGGAGCCAAAGTGCACCAAGAAAGAACCAATCCACCCAGATCTAGATC 180

QY 131 ArgSerProProGluAlaLeuValGlnGlyArgTyrProHisIleLysAspGlyGluAsp 150
 DB 181 AGGTCACTCTCTGAAGCCCTCGTTCAGGGTCGATATCCCAACATCAAGCAGGTGAGAT 240

QY 151 LeuLysAspHisSerThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHis 170
 DB 241 CTTAAAGACCATCTCAACAGAAAGTAAAAAATGGAATACTGTCTAGGAGAAATCCAGCAT 300

QY 171 GluValGluLysSerGluLeuSerGluAsnThrAspAlaSerGlyLysIleGluLysTyr 190
 DB 301 GAAGTAGAAAAATCAGAAATCAGTCAAAACACAGATGCTTCGGGCAAAATAGAGAAATAT 360

QY 191 AsnValProLeuAsnArgLeuLysMetPheGluLysGlyGluProThrGlnThrLys 210
 DB 361 AATGTTCCGCTCAACAGGCTTAAGATGATGTTTGAGAAAGTGAACCACTCAACTAAG 420

QY 211 IleLeuArgAlaGlnSerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSer 230
 DB 421 ATTCTCCGGGCCCAAGCGGAAGTCAAGTGAAGAGAGATCTCTGAAACCTCTATAAGGATCGA 480

QY 231 LeuAspLeuGluIleGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLys 250
 DB 481 CTAGATGACCTGGAAATAGGCCCGAGTGTGTCATCTCTCTACATTTGACTCGGAGAA 540

QY 251 AsnGluSerArgAsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArg 270
 DB 541 AATGAGAGTAGACCAATCTGGAATCTCCAGCCTCTCAGAAACCTCTATAAGGATCGA 600

QY 271 MetAlaLysTyrGlnAlaAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGlu 290
 DB 601 ATGCCCAAGTACCAGGAGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAATGAG 660

QY 291 LeuLysAlaSerGlyGlyGluIleLysIleHisLysMetGluGlnLysGluAsnValPro 310
 DB 661 CTGAAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGAGCAAAAGGAGATGTGCC 720

QY 311 ProGly-ProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAs 330
 DB 721 CCAGGTCCCTGAGTCTGCAATCACCATCAGGAGGGGAAAGAAATTTCCGCCAATCAGAA 780

QY 330 nSer-LeuAlaValArg-SerThrProAlaGlu-AspAspSerArgAspSer-GlnVal 349
 DB 781 TAGCCCTGGCAGTCCGTTTCCACCCCTCGCGAAATATGACTCCCGGAGTCTCCCGAGTTA 840

QY 349 YsSerGlu-ValGlnGln-ProValHisProLys-ProLeuSerProAsp-SerArgAl 367
 DB 841 AAGGAGAGGGTTCAAGAGGCTGGCATCCCAAGGCCCTTAAGTCCAGATTTCCAAAGGC 900

QY 367 aSerSerLeuSer---GluSerSerProProLysAlaMetLysLys---PheGlnAlaPr 385
 DB 901 CTCAGGCTTTTGAAGGTTCTCCCTCCCAAGCCAAATGAAAGAAAGTTTTCGGGACC 960

QY 385 o---AlaArgGluThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLe 404
 DB 961 CTGGAAGAAAAAACCTGC-----CGGGGAAT 987

QY 404 uAlaAsnGlnGlnValPhe 410
 DB 988 GGCACAAAAACAAGTCTTT 1006

RESULT 7
 BE410108 971 bp mRNA linear EST 21-JUL-2000
 LOCUS BE410108
 DEFINITION 601302288F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3637100 5',
 mRNA sequence.
 ACCESSION BE410108
 VERSION BE410108.1 GI:9346558
 KEYWORDS EST
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 971)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-f@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLC335 row: o column: 21
 High quality sequence start: 56
 High quality sequence stop: 806
 Location/Qualifiers
 1. 971
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3637100"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_21"
 /note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
 Site 2: EcoRI; cDNA made by oligo-dT priming
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGACGAG(G). Size-selected 500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 290 a 236 c 243 g 201 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 2.41e-90 Length: 971
 Score: 1294.00 Matches: 268
 Percent Similarity: 85.85% Conservative: 11
 Best Local Similarity: 82.45% Mismatches: 33
 Query Match: 32.95% Indels: 14
 DB: 10 Gaps: 6

US-09-890-549-4 (1-759) x BE410108 (1-971)

QY 269 AspArgMetAlaLysTyrGlnAlaValSerLysGlnSerSerThrAsnTyrThr 288
 DB 28 AATCGCGCAGCGCCAGTCCAGGAGCTGTGTCAAAACAAAGCAGCTCAACCACTATACA 87

QY 289 AsnGluLeuLysAlaSerGlyGluIleLysIleHisLysMetGluGlnLysGluAsn 308
 DB 88 AATGAGCTGAAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAGAGAAAT 417

QY 309 ValProProGluValCysIleThrHisGlnGluGlyGluValSerAlaAsn 328
 Db 148 GTCCCCCAGGTCTGAGGTCTGCATCACCATCAGAAAGGGAAGATTTCTGCAAT 207
 QY 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnVal 348
 Db 208 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAAGATGACTCCGCTGACTCCAGGTT 267
 QY 349 LysSerGluValGlnProValHisProLysProLeuSerProAspSerArgAlaSer 368
 Db 268 AAGAGTGGAGTTCACAGAGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 327
 QY 369 SerLeuSerGluSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388
 Db 328 AGTCTTCTGAAGTTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTCCAGAGAG 387
 QY 389 ThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsnGlnGln 408
 Db 388 ACCTGGGTGAATGTGAGAGACAGTCTATCCAATGGAGCGTCTCTTGCCCAACCCAGCAG 447
 QY 409 ValPheHisIleSerCysPheArgCysSerTyCysAsnLysLeuSerLeuGlyThr 428
 Db 448 GTGTTTTCATCATCAGTCTGCTTCCGTTCTCTATTGCAACAACAACCTCAGTCTAGGAACA 507
 QY 429 TyrAlaSerLeuHisGlyArgIleTyCysLysProHisPheAsnGlnLeuPheLysSer 448
 Db 508 TATGCATCTTTTACATGAAGAATCTATTGTAGCCTCACTTCAATCAACTCTTTAAATCT 567
 QY 449 LysGlyAsnTyAspGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLys 468
 Db 568 AAGGGCAACTATGATGAAGCTTTGGGCACAGACACACAGAGATCTATGGCAAGCAAAA 627
 QY 469 AsnGluAsnGluGluLeuLeuArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
 Db 628 AATGAAACGAAAGAGATTTTGAGAGAGACCCAGCTTGCAATGCAAGGGAGACCCCT 687
 QY 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaSer--- 507
 Db 688 CACAG-CCAGGGGTGAAGATCCCTCTATTGTAAGTGGGTGCTCTGGGTGCAAGTAT 746
 QY 508 MetGluAlaLysAlaSer-SerGlnGlnGluLysGluAspLysProAlaGluThrLysLy 527
 Db 747 TTGGAAGCCCAAGCTCTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 806
 QY 527 sleuArgIleAlaTrpProProProThrGluLeuGlySerSerGlySerAlaLeuGluGl 547
 Db 807 GCTGAAGATCGCTTGGCACCCCTCACTTGAAGTTCAGGA---GTGCTGGAGGA 863
 QY 547 uGlyIleLysMetSerLysProLysTrpProGluAspGluIleSerLysProGluVa 567
 Db 864 AGGGTCAAAATGTTAAC---CCCAATGCTCTGAAAGAGAA-----TAGCAGCCCGAGTT 914
 QY 567 lProGluAspValAspLeuLysLysLysLeuArgArgSerSerSerLeuLysGluAr 587
 Db 915 C-----CGGATGTCTTAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957
 QY 587 gSerArgProPhe 591
 Db 958 ---AAGCCCTTT 966
 RESULT 8
 BM552304
 LOCUS
 DEFINITION BM552304 989 bp mRNA linear EST 20-FEB-2002
 5', mRNA sequence.
 ACCESSION BM552304
 VERSION BM552304.1 GI:18790057
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 989)
 NIH-MGC http://mgi.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Cloning Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LiCM1994 row: d column: 22
 High quality sequence stop: 666.
 Location/Qualifiers
 1..989
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5476965"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_hosts="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_98"
 /note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 339 a 219 c 233 g 197 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,09e-87 Length: 989
 Score: 1252.50 Matches: 257
 Percent Similarity: 93.57% Conservative: 5
 Best Local Similarity: 91.79% Mismatches: 13
 Query Match: 31.89% Indels: 5
 DB: 12 Gaps: 3
 US-09-890-549-4 (1-759) x BM552304 (1-989)
 QY 56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLys 75
 Db 130 AACACCCGAAATCTCTCCAGACATTTAGAAAGGGACCTGACTGTGTTAAAGAGAGAG 189
 QY 76 TrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerThr 95
 Db 190 TGGAGAACCCAGGGCTGGGAGCAGAGTCTCACAGACTCTCTACGGAACAGCAGCACT 249
 QY 96 GluIleArgHisArgAlaAspHisProProAlaGluValThrSerHisAlaLysSerGly 115
 Db 250 GAGATTAGCACAGACAGACCATCTCTCTGTAAGTGACAGCCACCTGCTCTTCTGA 309
 QY 116 AlaLysAlaAspGlnGluGlnIleHisProArgSerArgLeuArgSerProGlu 135
 Db 310 GCCAAGCTGACCAAGAGAAACAATCCACCACCATCTAGACTCAGGTCTCCTCTGAA 369
 QY 136 AlaLeuValGlnGlyArgTyrProHisIleLysAspGlyGluAspLeuLysAspHis 155
 Db 370 GGCCTCGTTTCCAGGTGCGATATCCACATCAAGGACGGTGGAGATCTTAAAGACCACTCA 429
 QY 156 ThrGluSerLysLysMetGluAsnLysCysLeuGlyGluSerArgHisGluValGlyLys 175
 Db 430 ACAGAAAGTAAATAATGAAAAATTTCTAGAGAAATCCAGGCATCCAGGCAATGAAAGTCA 489
 QY 176 GluIleSerGluAsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsn 195
 Db 490 GAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAAC 549

QY 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGln 215
 Db 550 AGGCTTAAGATGATGTTTGAAGAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCAA 609
 QY 216 SerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSerIleuAspAspLeuGlu 235
 Db 610 AGCCGAGTCAAGTGAAGAGAGATCTCTGAAGACAGCTATTCTCTAGATGACCTGNA 669
 QY 236 IleGlyProGlnLeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArg 255
 Db 670 ATAGGCCAGGTGAGTTGTCTCTACATTTCATCGGAGAAAATAGAGTAGACGA 729
 QY 256 AsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGln 275
 Db 730 AATCTGGAACCTCCAGCCCTCTCAGAAACCTCTATTAAGATCGATGCCCAGATACCA 789
 QY 276 AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeu-LysAlaSerGln 295
 Db 790 GCAGCTGTGTCACAAACAAAGCAGCTCAACCCACTATACAAATGAGTGAAGAACCCGGGG 849
 QY 295 YGlyGluIleLysIleHisLysMet-GluGlnLysGlu--AsnValProProGlyProG 314
 Db 850 GGCGCAAAATCAAAATCTCTAAATGGAGCAAAAGGAAAGGTGGCCCGGAGGTCCGG 909
 QY 314 lu---ValCysIleThrHis---GlnGluGlyGluLysIleSerAlaAsnGluAsn 330
 Db 910 AAGGTCTGCTTACCCCATAGGAAAGGGGAAAATAATTTCTGGGCAAAAT 965

RESULT 9
 BG476496 768 bp mRNA linear EST 21-VAR-2001
 LOCUS 602522032F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4640502 5',
 DEFINITION mRNA sequence.
 ACCESSION BG476496
 VERSION BG476496.1 GI:13408775
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 768)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapb-remail.nih.gov
 Tissue Procurement: ATCC/DCTP
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LICM1405 row: p column: 07
 High quality sequence stop: 745.
 Location/Qualifiers
 1..768
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4640502"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_20"
 /notes="Organ: Skin; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the Laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 267 a 150 C 207 G 141 t 3 others

ORIGIN

Alignment Scores:

Pred. No.: 1..62e-86 Length: 768
 Score: 1243.00 Matches: 251
 Percent Similarity: 97.6% Conservative: 0
 Best Local Similarity: 97.6% Mismatches: 3
 Query Match: 31.65% Indels: 3
 DB: 10 Gaps: 0

US-09-890-549-4 (1-759) x BG476496 (1-768)

QY 504 LeuAlaLysSerMetGluAlaLysSerGlnGlnLysGluAspLysProAla 523
 Db 2 CTGGTGTCAAGTATGGAAGCCAGGCTCTCTCAGCAGGAGAGAAAGAACAGACGACCT 61
 QY 524 GluThrLysLysLeuArgIleAlaTyrProProThrGluLeuGlySerSerGlySer 543
 Db 62 GAAACCAAGAGCTGAGGATCGCTGGCCACCCCTGAACTTGAAGTTCAGGAAGT 121
 QY 544 AlaLeuGluGluGlyIleLysMetSerLysProLysTyrProGluAspGluLysSer 563
 Db 122 GCCTTGGAGGAGGATCAAAATGTCAAGCCCAATGGCTCTCTGAAGACGAATCAGC 181
 QY 564 LysProGluValProGluAspValAspLeuAspLysLysLysLeuArgSerSerSer 583
 Db 182 AAGCCCGAAGTCTCTGAGGATGCTAGATCTGAAGAGGCTAAGACCATCTCTTCA 241
 QY 584 LeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThrSerValLys 603
 Db 242 CTGAAGGAAAGAGCCGCCCATTCACCTGTAGAGCTTCATTTCAAGCACCTCTGTCAAG 301
 QY 604 SerProLysThrValSerProProIleArgLysGlyTyrSerMetSerGluGlnSerGlu 623
 Db 302 AGCCCAAAAACTGTGTCCCACTATCAGGAAAGGCTGGAGCATGTCAGAGCAGAGTAA 361
 QY 624 GluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLys 643
 Db 362 GAGTCTGTGGGTGGAAGAGTTCAGAGAAAGGAAACAAAGTGGAATGCCAAGGCTTCTAAG 421
 QY 644 LysAsnGlyAsnValGlyLysThrThrTyrGlnAsnLysGluSerLysGlyGluThrGly 663
 Db 422 AAGAATGGGAATGTGGAAAAACAACCTGGCAAAACAAGAATCTAAAGAGAGACAGGG 481
 QY 664 LysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGly 683
 Db 482 AAGAGAAGTAAGGAAGTTCATAGTNTGGAGATGGAGAATGAGAATCTTGAGAAAATGTG 541
 QY 684 AlaAspSerAspGluAspAspAsnSerPheLeuLysGlnGlnSerProGlnGluProLys 703
 Db 542 GCAGACTCCGATGAAGATGATAACAGCTTCTCAACAAACAATCTCCCAAGAACCCAG 601
 QY 704 SerLeuAsnTyrSerSerPheValAsp-AsnThrPheAlaGluGluPheThrThrGlnAs 723
 Db 602 TCTCTGAATNGTTCAGTTTNGTAGACAACACCTTTGCTGAAGAATTCACCTACTCAGAA 661
 QY 723 nGlnLysSerGlnAspValGluLeu-TyrGluGlyGluValValLysGluLeuSerValG 743
 Db 662 TCAGAAATCCCGAGGATGTGGAACTCTTGGAGCGGAGAGTGTGTCAAAGAGCTCTCTGTGG 721
 QY 743 luGluGlnIleLysArgAsnArgTyrTyrAspGlu-AspGluAsp 757
 Db 722 AAGAACAGATAAGAGAAATCGGTATTATGATGAAGGATGAGGAT 766

RESULT 10
 BU850203
 LOCUS BU850203
 DEFINITION AGNCOURT_10438215 NIH_MGC_109 Homo sapiens cDNA clone
 IMAGE:6598982 5', mRNA sequence.
 ACCESSION BU850203
 VERSION BU850203.1 GI:24035166
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 918)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM2832 row: k column: 14
 High quality sequence stop: 655.
 Location/Qualifiers
 1. 918
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="NIH_MGC_109"
 /lab_host="DH10B (phage-resistant)"
 /tissue_type="teratocarcinoma, cell line"
 /note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
 XhoI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH MGC Library."
 BASE COUNT 313 a 207 c 211 g 187 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 8.64e-85 Length: 918
 Score: 1222.00 Matches: 249
 Percent Similarity: 98.04% Conservative: 1
 Best Local Similarity: 97.65% Mismatches: 1
 Query Match: 31.12% Indels: 4
 DB: 13 Gaps: 0
 US-09-890-549-4 (1-759) x BU850203 (1-918)
 QY 56 AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys 75
 DB 156 AACACCGAAATCTTCCAGCAGCTTTAGAAAGGGGACCTGCTGTGTTAAAGAGAAG 215
 QY 76 TrpGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerThr 95
 DB 216 TGGGAGAACCCAGGGCTGGAGCAGAGTCTCACAGACTCTCTACGGAACAGCAGCACT 275
 QY 96 GluLeuArgHisArgAlaSerHisProAlaGluValThrSerHisAlaAlaSerGly 115
 DB 276 GAGATTAGCAGACAGACAGACCATCTCTCTGCTGAAGTGACAGCCAGCAGCTCTCTGA 335
 QY 116 AlaLysAlaAspGlnGluGlnLeuHisProArgSerArgLeuArgSerProGlu 135
 DB 336 GCCTAAGCTGACAGAGAGAAACAAATCCACCCAGATCTAGACTCAGTCCCTCTGAA 395
 QY 136 AlaLeuValGlnGlyArgTyrProHisLeuLysAspGlyGluAspLeuLysAspHisSer 155
 DB 396 GCCTCGTTTCAGGTCGATATCCCATCATCAGAGCGGTGAGGATCTTTAAAGACCACTCA 455
 QY 156 ThrGluSerLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSer 175
 DB 456 ACAGAAAGTAAAAAATGMAAATTTCTAGAGAGATCCAGGCATCAGAGTGAAGTGAATAATCA 515
 QY 176 GluLeuSerGluAsnThrAspAlaSerGlyLysLeuGluLysTyrAsnValProLeuAsn 195

Db 516 GAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAATATATATGTTCCGCTGAAC 575
 QY 196 ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysLysLeuArgAlaGln 215
 Db 576 AGCTTTAAGATGATGTTTGAGAAAGTGAAACCACTCAAACTAAGATTCTCCGGGCCCAA 635
 QY 216 SerArgSerAlaSerGlyArgLysLysSerGluAsnSerTyrSerLeuAspLeuGlu 235
 Db 636 AGCCGAAGTGCAAGTGAAGGAGATCTCTGAAAACAGCTATCTCTAGATGACCTGGAA 695
 QY 236 IleGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArg 255
 Db 696 ATAGGCCAGGTCAGTTGTCTTCTACATTTTGAAGGAGAAAATCAGAGTAGACGA 755
 QY 256 AsnLeuGluLeuProArgLeuSerGluThrSerLysLysAspArgMetAlaLysTyrGln 275
 Db 756 AATCTGGAACCTTCCAGGCTCTCAGAAACCTCTATAAAAGATCGAATGCGCAAGTACCAG 815
 QY 276 AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerG 295
 Db 816 GCAGCTGTGTCCCAACAAAGCAGCTCCACCAACTATCAAAATGAGTGGAAAGCAGTG 875
 QY 295 lyGly-GluLeuLysLysHis-LysMetGluGlnLys 306
 Db 876 GGGGGGAAATCAAAATTCATAAAATGGAGCCAAAG 912
 RESULT 11
 BU861039
 LOCUS AGENCOURT 10436477 NIH_MGC 126 Homo sapiens cDNA clone
 DEFINITION IMAGE:6652372 5', mRNA sequence.
 ACCESSION BU861039
 VERSION BU861039.1 GI:24046031
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 787)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: NCI
 CDNA Library Preparation: Michael Brownstein Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM2900 row: 1 column: 04
 High quality sequence stop: 617.
 Location/Qualifiers
 1. 787
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6652372"
 /tissue_type="mixed (pool of 40 RNAs)"
 /lab_host="DH10B (Ti-phage-resistant)"
 /clone_lib="NIH_MGC_126"
 /notes="Vector: pDNR-LIB; Site 1: SfiI (ggcattatggcc);
 Site 2: SfiI (ggccctccggcc); Double-stranded cDNA was
 prepared from a pool of 40 cell line polyA+ RNAs (bladder
 - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
 kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
 ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
 gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
 used in cloning as follows:
 5'-AAGCAGTGTATCAACGACAGTGGCCATTACGGCGGG-3' and

FEATURES
 source

QY 696 nGlnSerProGlnGluProLysSerLeuAsnTrpSerPheValAspAsnThrPheAl 716
 Db 545 ACAATCTCCACGAACCAAGTCTCTGAATGGTCGAGTTTGTAGACAACACCTTTGC 604
 QY 716 aGluGluPheThrThrGlnAsnGlnLysSerClnAspValGluLeuTrpGluGlyGluVa 736
 Db 605 TGAAGAAATCTACTACTCAGAAATCCAGGATGCGAATCTCGGAGGAGAGT 664
 QY 736 lValysGluLeuSerValGluGluGlnLysLeuArgAsnArgTyrTyrAspGluAspG1 756
 Db 665 GGTCAAGAGTCTCTGTGGAGAACCAAGATACGAGAAATCGGTATCATGATGAGGATGA 724
 QY 756 uAspGluGlu 759
 Db 725 GGATCAAGAG 734

RESULT 14
 LOCUS BI089829 770 bp mRNA linear EST 20-JUN-2001
 DEFINITION 602855072F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4996480 5',
 mRNA sequence.
 ACCESSION BI089829
 VERSION BI089829.1 GI:14508159
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 770)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM1021 row: p column: 17
 High quality sequence stop: 707.
 Location/Qualifiers
 1..770
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4996480"
 /cell_line="MGC36"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_10"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: Not I;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.5 kb. Library prepared by Life
 Technologies."
 BASE COUNT 250 a 194 c 174 g 152 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.85e-83 Length: 770
 Score: 1199.50 Matches: 245
 Percent Similarity: 96.51% Conservative: 4
 Best Local Similarity: 94.96% Mismatches: 5
 Query Match: 30.54% Indels: 6
 Db: 12 Gaps: 0

US-09-890-549-4 (1-759) x BI089829 (1-770)

QY 233 AspLeuGluLeuGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGlu 252
 Db 3 GACCTGGAATAGCCCAAGGTCAGTTGTCATCTTCTACATTTGACTCGGAGAAAAATGAG 62

QY 253 SerArgAsnLeuGluLeuProArgLeuSerGluThrSerIleLysAspArgMetAla 272
 Db 63 AGTAGAGCAAAATCTGGAACCTTCCACGCTCTCAGAAACCTCTATAAAGGATCGAATGCC 122
 QY 273 LysTyrGlnAlaAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeuLys 292
 Db 123 AAGTACAGGACGCTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAA 182
 QY 293 AlaSerGlyGlyGluLeuLysIleHisLysMetGluGlnLysGluAsnValProProGly 312
 Db 183 GCCAGTGTGGCGAATCAAAATTCATAAATGGACCAAAAGAGAAATGTGCCCCCAGGT 242
 QY 313 ProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeu 332
 Db 243 CCTGAGGTCTGCATCACCCTCAGGAAGGGGAAAAGATTCTGCAATAGGAATACCTTG 302
 QY 333 AlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnValLysSerGluVal 352
 Db 303 GCAGTCGGTTCACCCCTCCGGAAGTGAATCCCTGACTCCCGTAAAGAGTGAAGT 362
 QY 353 GlnGlnProValHisProLysProLeuSerProAspSerArgAspSerArgAspSerLeuSerGlu 372
 Db 363 CAACAGCCTGTCATCCCAAGCCCTAAGTCCAGATACCAAGAGCCTCCAGTCTTTCTGAA 422
 QY 373 SerSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGlu 392
 Db 423 AGTCTCTCTCCCAAGCAATGAAGATTACAGGCACCTGCAAGAGAGACCTGCGTGGAA 482
 QY 393 CysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIle 412
 Db 483 TGTCAGAGACAGTCTATCCAATGAGCGTCTCTTGGCCCAACCAGCAGGTGTAGCACATC 542
 QY 413 SerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeu 432
 Db 543 AGCTCTCTCCGTGTCTCTATTGCAACCAACAACTCAGTCTAGGAACATATGTCATCTTA 602
 QY 433 HisGlyArg--IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnT 452
 Db 603 CATGGACAGCAATCTATCGTAAGCCTCACTTCAATCAACTCTTTAAATCTTAAGGCACT 662
 QY 452 TyrAsp-GluGlyPheGlyHis-ArgProHisLysAspLeuTrpAlaSerLysAsnGluAs 471
 Db 663 ATGATGCAAGGCTTTGGCCACCACACCAAGATCTATGGCAAGCAAGAAAAATGAAA 722
 QY 471 nGluGluLeuGluArgProAlaGlnLeuAlaAsnAlaArgGlu 486
 Db 723 CGAAGAGATAGC-GAGAGACGAGCCAGAGA-GCCAATGCAAGGGAG 766

RESULT 15
 LOCUS BM449793 931 bp mRNA linear EST 05-FEB-2002
 DEFINITION AGENCOURT_6392972 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528511
 5', mRNA sequence.
 ACCESSION BM449793
 VERSION BM449793.1 GI:18498833
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 931)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/PTP
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LLAM12205 row: h column: 16
High quality sequence stop: 575.
Location/Qualifiers
1. .931
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:528511"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 296 a 267 c 192 g 176 t
ORIGIN

Alignment Scores:
Pred. No.: 1.56e-82 Length: 931
Score: 1193.00 Matches: 246
Percent Similarity: 91.64% Conservative: 6
Best Local Similarity: 89.45% Mismatches: 18
Query Match: 30.38% Indels: 5
DB: 12 Gaps: 2
US-09-890-549-4 (1-759) x BM449793 (1-931)

Qy	119	AspGlnGluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuVal	138
Db	1	GACCAAGAGAACAAATCCACCCAGATCTAGACTCAGGTCCACCTCTGAGCCCTCGTT	60
Qy	139	GlnGlyArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisThrGluSer	158
Db	61	CAGGGTCGATATCCCAATCAAGACGGTGAGGATCTTAAGACCACTCAACAGAAAGT	120
Qy	159	LysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSer	178
Db	121	AAAAAATGGAATTCCTAGGAGATCCAGGATGAGTGAAGTGAAGTGAAGTGAAGT	180
Qy	179	GluAsnThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLys	198
Db	181	GAAACACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGACAGGCTTAAG	240
Qy	199	MetMetPheGlyLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSer	218
Db	241	ATGATGTTTGAGAAAGGTGAACCACTCAAACTAAGATTCTCCGGGCCCAAGCCGAGT	300
Qy	219	AlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyPro	238
Db	301	GCAAGTGAAGGAGATCTCTGAAACACACTATTCTTAGATGACCTGGAAATAGGCCCA	360
Qy	239	GlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGlu	258
Db	361	GGTCAGTTGTCATCTCTACATTTGACTCGAGAGAAATGAGAGTAGACGAAATCTGGAA	420
Qy	259	LeuProArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaVal	278
Db	421	CTTCCACGCCCTCTCAAAACCTCTATAAGAGTCAATGGCCCACTACCAGGAGCTGTG	480
Qy	279	SerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGluIle	298
Db	481	TCCAAACAAAGCAGCTCAACCACTATCAAAATGAGCTGAAAGCCAGTGTGGGAAATC	540
Qy	299	LysIleHisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThr	318
Db	541	AAATTCATAAAATGAGCAAAAGGAGATGTGCCCCCAGGTCTCTGAGGTCTGCATCACC	600
Qy	319	HisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrPro	338
Db	601	CATCAGGAAGCGAAAGATTCTGCATCAGAAATAGCCTGGCAGGCCCTCCCTCCCTCCCT	660

Qy	339	AlaGluAspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisPro	358
Db	661	GCCGAAAGATGATTTCCGCTGACTTCTATGTTACCACTGAGGTTCACACAGCTGCTCCCTCCC	720
Qy	359	Lys-ProLeuSerProAspSerArgAla---SerSerLeuSerGluSerSerPro--Pro	376
Db	721	AACGCAACTCAGTCCAGACCCCGAGCCCCCCCCCTTTCCCGAAAAAGCTCTCCCTCCCA	780
Qy	377	LysAlaMetLysLysPhe---GlnAlaProAlaArgGlu	388
Db	781	AGCGAACGGAAAAAATTCCTCAGGGCACCCGCAAAAAAAA	819

Search completed: January 7, 2004, 01:45:14
Job time : 3947 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 6, 2004, 22:10:26 ; Search time 110 Seconds
(without alignments)
3045.544 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 3927

Sequence: 1 MESSFPNRQWTSLSLRVTA.....LSVERQIKRNYDEDEEE 759

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame, p2n.model -DEV=xlh

-O=/cgn2_1/USPTO_spool/US09890549/runat_06012004_094752_19841/app_query.fasta_1.903

-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCI=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -NATRIX=blosum62 -TRANS=numa40.cdi

-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US09890549 @CGN_1_1_56@runat_06012004_094752_19841 -NCPU=6 -ICPU=3

-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:

2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:

3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:

4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:

5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:

6: /cgn2_6/ptodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3927	100.0	2905	4	US-09-620-312D-45
2	606	15.4	413	4	US-09-016-434-680
3	442.5	11.3	710	4	US-09-016-434-701
4	392	10.0	1638	4	US-09-484-970B-20
5	374.5	9.5	654	4	US-09-016-434-994
6	368	9.4	238	4	US-09-016-434-181
7	226.5	5.8	6755	3	US-08-911-999-4
8	226	5.8	4226	4	US-09-620-312D-480
9	223	5.7	6002	4	US-09-345-882-4
10	212.5	5.4	988	4	US-09-282-146-1
11	205.5	5.2	1015	3	US-08-934-627B-1
12	196	5.0	4323	4	US-09-620-312D-134

13	195.5	5.0	5857	4	US-09-220-132-79	Sequence 79, Appl
14	194	4.9	8532	1	US-08-452-655B-1	Sequence 1, Appl
15	194	4.9	8532	3	US-08-450-582-1	Sequence 1, Appl
16	194	4.9	9606	1	US-07-741-940-1	Sequence 1, Appl
17	194	4.9	9606	1	US-08-289-548A-1	Sequence 1, Appl
18	194	4.9	9606	1	US-08-452-654-1	Sequence 1, Appl
19	194	4.9	9606	2	US-08-370-215A-1	Sequence 1, Appl
20	194	4.9	9606	4	US-08-449-731-1	Sequence 1, Appl
21	193	4.9	11236	1	US-07-853-913-1	Sequence 1, Appl
22	192	4.9	6605	1	US-08-769-309A-4	Sequence 1, Appl
23	192	4.9	6605	3	US-08-994-570-4	Sequence 4, Appl
24	192	4.9	6608	4	US-08-220-132-58	Sequence 4, Appl
25	191	4.9	7453	4	US-09-620-312D-248	Sequence 58, Appl
26	191	4.9	7501	4	US-09-620-312D-249	Sequence 249, App
27	190	4.8	6201	2	US-08-790-912-1	Sequence 1, Appl
28	187.5	4.8	11907	3	US-08-061-376-4	Sequence 1, Appl
29	187.5	4.8	14255	1	US-08-320-559-1	Sequence 1, Appl
30	187.5	4.8	14255	1	US-08-327-392-1	Sequence 1, Appl
31	187.5	4.8	14255	1	US-08-306-691B-55	Sequence 55, Appl
32	187.5	4.8	14255	3	US-08-545-860D-1	Sequence 1, Appl
33	187.5	4.8	14255	5	PCT-US94-04496-1	Sequence 1, Appl
34	185.5	4.7	580073	4	US-08-545-528D-1	Sequence 1, Appl
35	184.5	4.7	3115	4	US-09-620-312D-58	Sequence 58, Appl
36	180	4.6	3972	4	US-09-645-456A-6	Sequence 6, Appl
37	180	4.6	3972	4	US-09-425-324A-6	Sequence 6, Appl
38	180	4.6	3972	4	US-09-645-791-6	Sequence 6, Appl
39	179.5	4.6	4967	3	US-09-177-325-12	Sequence 12, Appl
40	179.5	4.6	4967	3	US-09-411-812A-12	Sequence 12, Appl
41	179.5	4.6	4967	4	US-09-590-113-12	Sequence 12, Appl
42	179.5	4.6	6744	1	US-08-119-125A-2	Sequence 2, Appl
43	179	4.6	15213	4	US-08-961-527-26	Sequence 26, Appl
44	178.5	4.5	4215	4	US-09-620-312D-295	Sequence 295, App
45	178	4.5	3807	4	US-09-645-456A-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1

US-09-620-312D-45
; Sequence 45, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aigong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE OF INVENTION: Polypeptides
; FILE REFERENCE: 78CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 45
; LENGTH: 2905
; TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (173)...(2452)
US-09-620-312D-45

Alignment Scores:

Pred. No.: 0 Length: 2905
Score: 3927.00 Matches: 759
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-890-549-4 (1-759) x US-09-620-312D-45 (1-2905)

QY 1 MetGluSerSerProPheAsnArgArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
DB 173 ATGGAATCATCTCCATTAAATAGACGGCAATGGACCTCACTATCATTTAGGGTAAACAGCC 232
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
DB 233 AAAGAACTTCTCTGTGCAACAAGAACAGTCAATCGGCTATTGTGGAATATTCTCCAG 292
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
DB 293 TACCAGAAACAGCTGAGAAACAAACATCGAGAGAGAGAGAGTAAACCCGAAATCTC 352
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
DB 353 TCCAGCACATTTAGAAAGGGGACCTGACTGTGTAAAGAGAGAGTGGAGAAACCCAGGG 412
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
DB 413 CTGGAGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACTGAGATTAGGCACAGA 472
QY 101 AlaAspHisProAlaGluValThrSerHisAlaSerGlyAlaLysAlaAspGln 120
DB 473 GCAGACCATCTCTCTGCTGAGTACACAGCCGCTGCTTCTGGAGCCAAAGCTGACCAA 532
QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProGluAlaLeuValGlnGly 140
DB 533 GAAGAAACAAATCCACCCAGATCTAGACTCAGTCACTCTCTGAGCCCTCTCTTCAAGGT 592
QY 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
DB 593 CGATATCCCCACATCAAGACCGTGGAGTCTTAAAGACCCTCAACAGAAAGTAAAGAA 652
QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
DB 653 ATGGAAATTTGCTAGGAGAAATCCAGGCATGAAGTAGAAGAAATCAGAAATCAGTGAAAC 712
QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
DB 713 ACAGATGCTTCGGGCAAAATACAGAAATATATATGTTCTCGTGAAACAGGCTTAAAGATG 772
QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
DB 773 TTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAGTGCAGT 832
QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
DB 833 GGAAGGAGATCTCTGAAACACGCTATTCTCTAGATGACCTGGAATAGGCCAGGTCAG 892
QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
DB 893 TTGTCATCTTCTACATTTGACTCGAGAGAAATAGAGTAGACGAAATCTGGAACTTCCA 952
QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
DB 953 CGCTCTCAGAAACCTCTATAAGGATCGAATGCCAAGTACGAGGAGCTGTGTCCAAA 1012
QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300

DB 1013 CAAAGCAGCTCAACCAATATACAAATGAGCTGAAAGCCAGTGGTGGCGAAATCAAAAT 1072
QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
DB 1073 CATAAATCGAGCAAAAGAGAAATGTGCCCCAGGTCTCTGAGGTCTGCATCATCCCATCAG 1132
QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
DB 1133 GAAGGGGAAAGATTTCTGCAATAGAAATAGCTGGCAGTCCGTTCACCCCTCCCGAA 1192
QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
DB 1193 GATGACTCCCGTACTCCAGGTTAAGAGTGAAGTTCACAGCCTGTCTCATCCCAAGCCA 1252
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
DB 1253 CTAAGTCCAGATTCAGAGACCTCCAGTCTTTCTGAAAGTTCTCTCCCCAAAGCAATGAAG 1312
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
DB 1313 AAGTTTCAGGCACCTCGCAAGAGAGACCTGCGTGAATGTCAAGACAGTCTATCCAATG 1372
QY 401 GluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
DB 1373 GAGCGTCTCTGGCCCAACACAGCAGGTGTTTCACATCAGCTGCTTCCTGCTCTTATTC 1432
QY 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
DB 1433 AACCAACAACTCAGTCTAGGAACATATGCATCTTTACATGGAAGAACTATTTGAAGCCT 1492
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
DB 1493 CACTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGCTTTGGCAGACAGACCA 1552
QY 461 HisLysAspLeuTyrAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAlaGln 480
DB 1553 CACAAGGATCTATGGGCAAGCAAAATGAAACGAAGAGATTTTGGAGAGACCAAGCCAG 1612
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
DB 1613 CTTGCAATGCAAGGAGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTTATTTGCTAAG 1672
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysLysGluAsp 520
DB 1673 GTGGGTCTCTGGTGTCAAGTATGGAAGCAAGCCCTCTCTCAGCAGAGAGAGAGAGAC 1732
QY 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProProThrGluLeuGlySer 540
DB 1733 AGCCAGCTCAAAACCAAGAGAGCTGAGGATCGCTGCGCCACCCCTCAACTGGAAGT 1792
QY 541 SerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTrpProProGluAsp 560
DB 1793 TCAGGAAGTGCCTTGGAGGAGAGGATCAAAATGTCAAGCCCAATGGCTCTCTGAAGAC 1852
QY 561 GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
DB 1853 GAAATCAGCAAGCCCAAGTCTCTGAGGATGTCATCTAGATCTGAAGAGCTTAAGACGA 1912
QY 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
DB 1913 TCTTCTTCTACTGAAGGAAAGAGAGCCGCCATTCCTGTAGCAGCTTCAATTTCAAAGCACC 1972
QY 601 SerValLysSerProLysThrValSerProProIleArgLysGlyTrpSerMetSerGlu 620
DB 1973 TCTGTCAAGAGCCCAAACTGTGTCCCTATCAGGAAAGGCTGGAGCATGTCCAGAG 2032
QY 621 GlnSerGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
DB 2033 CAGAGTGAAGAGTCTCTGGTCAAGAGTTCAGAAAGGAAACAAGTGGAAATGCCAAG 2092
QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGly 660

Db 2093 GCTTCTAAGAAGATGGGATGGGAAACAACTGGCAAAACAAAGAAATCTAAAGGA 2152
QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 2153 GAGACAGGGAAGAGAGTAAAGAGGTCTAGTTTGGAGATGGAGATGAGAAATCTTGA 2212
QY 681 GluAsnGlyAlaAspSerLysAspGluAspAspSerPheLeuLysGlnGlnSerProGln 700
Db 2213 GAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCTCAACAACTCTCCAA 2272
QY 701 GluProLysSerLeuAsnTrpSerPheValAspAspThrPheAlaGluGluPheThr 720
Db 2273 GAACCCAGTCTCTGAATGGTCGAGTTTGTAGACAAACACACTTTGCTGAGAAATCTACT 2332
QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTyrGluGluValValLysGluLeu 740
Db 2333 ACTCAGATCAGAAATCCAGAGTGTGAGACTCTGGAGCGGAGAGTGGTCAAAGAGCTC 2392
QY 741 SerValGluGluGlnLysArgAsnArgTyrTyrAspGluAspGluGluGlu 759
Db 2393 TCTGTGGAGAACAGATAAGAGAAATCGGTATTATGATGAGATGAGATGAGAG 2449
RESULT 2
US-09-016-434-680
; Sequence 680, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016.434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 680:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 413 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2NOT01
; CLONE: 269891
US-09-016-434-680
Alignment Scores: 2.16e-50 Length: 413
Pred. No.: 606.00 Matches: 127
Score:

Percent Similarity: 92.09% Conservative: 1
Best Local Similarity: 91.37% Mismatches: 8
Query Match: 15.43% Indels: 4
DB: 4 Gaps: 0

US-09-890-549-4 (1-759) x US-09-016-434-680 (1-413)

QY 296 GlyGluLeuLysLysLysLysMetGlu-GlnLysGluAsnValProGlyProGluVal 315
Db 3 GGCAGAAATCAAAATTCATAAAATGGAGNCAAAAGAGAAATGTGCCCCCAGGTCTGAGGT 62
QY 315 LCVSILGTHHISGLN-GluGlyLysLysLysSerAlaAsnGluAsnSerLeuAlaVal 335
Db 63 CTGCATCACCATCAGGAAAGGGGAAAGATTTCTGCAAAATAG-AAATAGCTGGCAGTCC 121
QY 335 rGserThrProAlaGluAspSerArgAsp-SerGlnValLysSerGluValGlnGln 354
Db 122 GTTCCANCCCTGCGAAGATGACGGGGGGAATCTCCANTTTNTGAGTGAGGTCAACAG 181
QY 355 ProValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSer 374
Db 182 CCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCTCCAGTCTTCTGAAAGTTCT 241
QY 375 ProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGln 394
Db 242 CCTCCCAAGCAATGAAGAGTTTTCAGGCACCTGCAAGAGAGACCTGCGTGGATGTGAG 301
QY 395 LysThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHisLysSerCys 414
Db 302 AAGACAGTCTATCCATGGAGCGTCTCTTGGCCAAACAGCAGGTGTTTACATCAGTGC 361
QY 415 PheArgCysSerTyrCysAsnLysLeuSerLeuGlyThrTyrAlaSer 431
Db 362 TTCCGTGCTCTATTGCAACAACTAGTCTAGGAACATATGCATCT 412

RESULT 3

US-09-016-434-701
; Sequence 701, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016.434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 701:

SEQUENCE CHARACTERISTICS:
LENGTH: 710 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HEARNOT01
CLONE: 305403
US-09-016-434-701

Alignment Scores:
Pred. No.: 5.18e-34 Length: 710
Score: 442.50 Matches: 101
Percent Similarity: 53.06% Conservative: 29
Best Local Similarity: 41.22% Mismatches: 88
Query Match: 11.27% Indels: 27
DB: 4 Gaps: 5

US-09-890-549-4 (1-759) x US-09-016-434-701 (1-710)

```
QY 387 ArgGluThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsn 406
Db 6 AGGAAATATGTCATCTTGTGCAAAAGRCAGTTTATCCAAATGGAGTGCCTAGTGGCAGAC 65
QY 407 GlnGlnValPheHisIleSerCysPheArgCysSerTyrcysAsnAsnLysLeuSerLeu 426
Db 66 AAGCAGAAATTTTCATAGTCTGCTCCGATGCCACCATTTGCAACAKTAATAAGTTTG 125
QY 427 Gly-ThrTyAlaSerLeuHisGlyArgIleTyrcysLysProHisPheAsnGlnLeuPh 446
Db 126 GGTAAATATGTCATCTTGTGCAAAAGRCAGTTTATCCAAATGGAGTGCCTAGTGGCAGAC 185
QY 446 eLysSerLysGlyAsnTyrcysLysPheGlyHisArgProHisLysAspLeuTrpAl 466
Db 186 CAAATCCAAAGGAAATATATGATGAAGTTTGTGACATAAGCAGCATAAAGATAGTGGAA 245
QY 466 aSerLysAsnGluAsnGluGluLys-----LeuGluArgProAlaGlnLeuAl 482
Db 246 CTCGAAATATGTCATCTTGTGCAAAAGRCAGTTTATCCAAATGGAGTGCCTAGTGGCAGAC 305
QY 482 aAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLysValGl 502
Db 306 AAATATTGCGAGAAACACCCCTTGTACCTGGAGATCGTAAATGACATTTA----- 354
QY 502 yValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp----- 520
Db 355 -----GATCGTGTGTAACAGTGAAGGCAAGGCAAGGAAATGATTGAG 392
QY 521 -LysProAlaGluThrLysLysLeuArgIleAlaLanProProProThrGluLeuGlySe 540
Db 393 AAAATTAGGGGAAAGCGGAAATTTAAAGTCAATTTGGCCCTCTCTCCAAAGGAGATCCCTAA 452
QY 540 rSerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTyrProProGlu-- 559
Db 453 GAAACCTTTACCTTTGAGGAAAGAGCTCAAAATGAGTAAACCTAAGTGGCCACCTGAAAT 512
QY 560 -AspGluIleSerLysProGluValProGluAspValAspLeu-AspLeuLysLysLeuA 579
Db 513 GACAACTTCCTATCCCTGAAATTTAAAGTCAATTTGCTAGGAGATGTTAGGACTCC 572
QY 579 rgArgSerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlns 599
Db 573 AGCAATAAAGNCNCAAGCAAGGANCACCTTTCCATTTTTCGACCTTATCTNCCAGTC 632
QY 599 erThrSerVal-----LysSerProLysThrValSerProProI 612
Db 633 CACCCCATGTTTGNCGAGGAGNGGTTGTTTNGGGTCCAAAGGAAATGGAATAGCCCT 692
QY 612 leArgLysGly 615
Db 693 TAAGGGGAGGG 703
```

RESULT 4

US-09-484-970B-20
Sequence 20, Application US/09484970B
Patent No. 6426186
GENERAL INFORMATION:
APPLICANT: Jones, Karen A.
APPLICANT: Volkmut, Wayne
APPLICANT: Walker, Michael G.
TITLE OF INVENTION: BONE REMODELING GENES
FILE REFERENCE: PB-0014 US
CURRENT APPLICATION NUMBER: US/09/484,970B
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PERL Program
SEQ ID NO 20
LENGTH: 1638
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6426186 200145-SCBI
LOCATION: 1554, 1581, 1624
OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-20

Alignment Scores:
Pred. No.: 1.81e-28 Length: 1638
Score: 392.00 Matches: 124
Percent Similarity: 42.60% Conservative: 43
Best Local Similarity: 31.63% Mismatches: 135
Query Match: 9.98% Indels: 93
DB: 4 Gaps: 13

US-09-890-549-4 (1-759) x US-09-484-970B-20 (1-1638)

```
QY 330 AsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnValLys 349
Db 103 AACCAGCGGGTGGCGCTTCTCCACCC-----GAGGCTTCCACTCCACAGCGCATGT 156
QY 350 SerGluValGlnGlnProValHisProLysProLeu---SerProAspSerArgAlaSer 368
Db 157 TCCAGGCTGAGGAGCGCCCGCAGCGCCACCCCTCTCATGAGCCCAAGGGGGGGGAGC 216
QY 369 SerLeuSerGluSerSerProProLysAlaMetLysLysPhe-----GlnAlaProAla 386
Db 217 AGCAGGTGAGCGCTCC-----AAGTCCTTACGCTGCGGGCCAGGTG 261
QY 387 ArgGluThrCysValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsn 406
Db 262 AAGGAGACCTTGGCGCGCTGCCAGAGACCGTGTACCCCATGGAGCGGTGGTGGCGGAC 321
QY 407 GlnGlnValPheHisIleSerCysPheArgCysSerTyrcysAsnAsnLysLeuSerLeu 426
Db 322 AAGCTCATTTTCCACAACTCTTGTCTGTCTGCAAGCACTGTCCACCAAGCTCAGCCTG 381
QY 427 GlyThrTyAlaSerLeuHisGlyArgIleTyrcysLysProHisPheAsnGlnLeuPhe 446
Db 382 GGAGCTACCGCGCTGCGCGGGAGTCTTACTGCAAAACCCCACTTCCAGCAGCTGTTT 441
QY 447 LysSerLysGlyAsnTyrcysLysPheGlyHisArgProHisLysAspLeuTrpAla 466
Db 442 AAGAGCAAAAGGCAACTTACGAGCGGGTTTGGCCGCAAGCAGCACAAGAGCTCTGGGCC 501
QY 467 SerLysAsnGlnGluLysGluAlaGluArgProAlaGlnLeuAlaAsnAlaArgGlu 486
Db 502 CACAAGGAGGTGAC----- 516
QY 487 ThrProHisSerProGlyValGluAspAla-----Profile 498
Db 517 -----CCCGGCACCAAGAGCGCTTGTAACTTCCACCCCTCTG 564
QY 499 AlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLys 518
Db ----- 518
```

RESULT 6
US-09-016-434-181

US-09-890-549-4 (1-759) X US-09-016-434-181 (1-238)

RESULT 7

US-09-890-549-4 (1-759) x US-08-931-999-4 (1-6755)

Qy	42	GlnLysAlaAlaGluThrAsnMetGluLysArgSerAsnThrGluAsnLeuSer	61
		:::::	
		:::::	
		:::::	
Db	3150	AGGAAAAGAAGGGGAACACAAGAAACGAAAAACAAACCGGCCAACAGCCCAAAACAAACGC	3209
		:::::	
		:::::	
Qy	62	GlnHisPheArg-----LysGlyThrLeuThrValLeuLys-----	73
		:::::	
		:::::	
Db	3210	GCCAGGGAAGACAGAGCGAAAGGGCAGAGGCGCACCAACACAAAGAAACAAAGCAA	3269
		:::::	
Qy	74	-----LysLysTrpGluAsnProGlyLeuGlyValaGluSerHisThrAspSer	89
		:::::	
		:::::	
Db	3270	AGCCGCACCCAAAAGAGGAGNACCCACCGAAAGGCGCAGCACAGAGACGAACACG	3329
		:::::	
Qy	90	LeuArgAsnSerSerThrGluIleArgHisArgAlaAspHisProAlaGluValThr	109
		:::::	
		:::::	
Db	3330	AGAAAAGACCAACAAACGAGCCACGACGAAAGACGAGAAAAGGAGGACACG-----	3383
		:::::	
		:::::	


```

Db      5234 AAAAACCGGAACGAGAAAGCAGAAACCAGAAAAGAGGAGCGAAAAAGAAAACGGC 5293 :
|||:::||| ||| ||| ::|||:::
Qy      758 luGlU 759
      ::::
Db      5294 AAAAG 5298
RESULT 8
US-09-620-312D-480
; Sequence 480, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A..
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillgrahast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 480
; LENGTH: 4226
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2982)
US-09-620-312D-480
Alignment Scores:
Pred. No.: 1.5e-11 Length: 4226
Score: 226.00 Matches: 181
Percent Similarity: 35.39% Conservative: 146
Best Local Similarity: 19.55% Mismatches: 319
Query Match: 5.76% Indels: 278
DBs: 4 Gaps: 43
US-09-890-549-4 (1-759) x US-09-620-312D-480 (1-4226)
QY      40 LysTyrGlnLysAlaAaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsn 59
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Ddb     250 GAATTCAGATGCCATTCCGAGAAAAGTTGTTAACAAGCAATGTAAGGAGTGATAAAT 309
QY      60 LeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnPro 79
      :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Ddb     310 GTAAAGAA-----ATAAAGTTAAGGAGGAAAATGAA 342
QY      80 GlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHis 99
      :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Ddb     343 ACAGAGATCCAGAAGAAATAAGATGCAGGAGGAGAGGAAT-----ATAATACCA 390
QY      100 ArgAlaAspHisProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAsp 119
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```
FEATURE:
NAME/KEY: allele
LOCATION: 1107..1125
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 1107..1125
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 1315..1338
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 1315..1338
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 1921..1967
OTHER INFORMATION: polymorphic fragment 5-136-174 SEQ ID41
FEATURE:
NAME/KEY: allele
LOCATION: 1921..1967
OTHER INFORMATION: polymorphic fragment 5-136-174 SEQ ID62
FEATURE:
NAME/KEY: allele
LOCATION: 3306..3352
OTHER INFORMATION: polymorphic fragment 5-143-84 SEQ ID46
FEATURE:
NAME/KEY: allele
LOCATION: 3306..3352
OTHER INFORMATION: polymorphic fragment 5-143-84 SEQ ID67
FEATURE:
NAME/KEY: allele
LOCATION: 1296..1338
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 1296..1338
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 3323..3369
OTHER INFORMATION: polymorphic fragment 5-143-101 SEQ ID45
FEATURE:
NAME/KEY: allele
LOCATION: 3323..3369
OTHER INFORMATION: polymorphic fragment 5-143-101 SEQ ID66
FEATURE:
NAME/KEY: allele
LOCATION: 4559..4605
OTHER INFORMATION: polymorphic fragment 5-148-352 SEQ ID48
FEATURE:
NAME/KEY: allele
LOCATION: 4559..4605
OTHER INFORMATION: polymorphic fragment 5-148-352 SEQ ID69
FEATURE:
NAME/KEY: misc_feature
LOCATION: 442..444
OTHER INFORMATION: ATG
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4378..4380
OTHER INFORMATION: stop : TGA
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 4878..4883
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 5116..5121
OTHER INFORMATION: potential
FEATURE:
```

```
NAME/KEY: polyA_signal
LOCATION: 5896..5901
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 5981..5986
FEATURE:
NAME/KEY: misc_feature
LOCATION: 209..756
OTHER INFORMATION: homology with EST in ref embl:W84531
FEATURE:
NAME/KEY: misc_feature
LOCATION: 391..815
OTHER INFORMATION: complement homology with EST in ref embl:W37603
FEATURE:
NAME/KEY: misc_feature
LOCATION: 453..898
OTHER INFORMATION: complement homology with EST in ref embl:H39516
FEATURE:
NAME/KEY: misc_feature
LOCATION: 818..1306
OTHER INFORMATION: complement homology with EST in ref embl:W67770
FEATURE:
NAME/KEY: misc_feature
LOCATION: 844..1303
OTHER INFORMATION: complement homology with EST in ref embl:AA262427
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1351..1702
OTHER INFORMATION: complement homology with EST in ref embl:AA485189
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1866..2109
OTHER INFORMATION: homology with EST in ref embl:AA296993
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2181..2281
OTHER INFORMATION: homology with EST in ref embl:T61718
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2253..2482
OTHER INFORMATION: homology with EST in ref embl:AA082927
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2480..2842
OTHER INFORMATION: complement homology with EST in ref embl:H38607
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3334..3733
OTHER INFORMATION: homology with EST in ref embl:AA279595
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3631..3870
OTHER INFORMATION: complement homology with EST in ref embl:AA169631
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3883..4221
OTHER INFORMATION: homology with EST in ref embl:H08612
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4277..4796
OTHER INFORMATION: homology with EST in ref embl:AA399016
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4516..5016
OTHER INFORMATION: homology with EST in ref embl:AA479433
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5580..6002
OTHER INFORMATION: complement homology with EST in ref embl:AA167428
US-09-345-882-4
```

Alignment Scores:

Pred. No.: 5,31e-11 Length: 6002
Score: 223.00 Matches: 178
Percent Similarity: 35.51% Conservative: 148
Best Local Similarity: 19.39% Mismatches: 326
Query Match: 5.68% Indels: 266
DB: 4 Gaps: 42

US-09-890-549-4 (1-759) x US-09-345-882-4 (1-6002)

QY 40 LysTyrGlnLysAlaAlaGluThrAsnMetGluLysLysArgSerAsnThrGluAsn 59
DB 1648 GAATTTGAGTGGATTGCCAGAGAAAGTTGTTACNAGCAATGTAAGGAGTGTGAAAT 1707

QY 60 LeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTyrGluAsnPro 79
DB 1708 GTAAAGAA-----ATAAAGTTAAGGAGGAATGAA 1740

QY 80 GlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluLeuArgHis 99
DB 1741 ACAGAGATCAAGAAATAAAGATGGAGGAGAGGAAT-----ATAATACCA 1788

QY 100 ArgAlaAspHisProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAsp 119
DB 1789 AGAAGAAAGCTATTGAGGAT-----GAAATTGAA 1821

QY 120 GlnGluGlnGlnHisPro-----ArgSerArgLeuArgSerProPro 134
DB 1822 AGAAAGAAATATTAGCCCTCTCTGGGAAGTAAAGAAAGATTTATTAGATCTATACCT 1881

QY 135 GluAlaLeuValGlnGlyArgTyrProHisIleLys-----AspGlyGluAspLeu 151
DB 1882 ACACATCTCTGATCAGGAAAGAAAGATTAACATTAAGAAACCAAGACATGAAATCTG 1941

QY 152 LysAspHisSerThrGluSerLysLysMetGluAsnCysLeu-----165
DB 1942 GATGACAAAGATGATGACACAACTAGGGTAGATGAAATCCCTCAACATAAGGTAGAACT 2001

QY 165 -----165

DB 2002 GAGGAGAAAGAAATCTCGAGATGAACGAATAAAGAAAGATGAAGATGATGAA 2061

QY 166 ---GlyGluSerArgHisGluValGluLysSerGluIleSerGluAsnThrAspAlaSer 184
DB 2062 GAAGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2121

QY 185 GlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMetPheGluLysGly 204
DB 2122 AATGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2181

QY 205 GluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer-----220
DB 2182 CGAGGAAAGAAATCAAAAAATGATGAAGCTAGTATTAAAGATTTCTGATGTTGAAGTGA 2241

QY 221 -----GlyA-glyLysIleSerGluAsnSerTyrSerLeu 231
DB 2242 GAGGTCCTTTACTTGTGTCATTCTCGGATGGAATGTGAGATACGATGAATGGATTAA 2301

QY 232 AspAspLeuGluIleGlyProGlyGlnLeuSer-----242
DB 2302 GCAGATAAAATAGTAAGACCTCTCATATAAATGTGCCAAGATAAATACATCGGAGAAA 2361

QY 243 ---SerSerThrPheAspSerGluLysAsnGlu-----SerArgArgAsnLeu 257
DB 2362 ATAAAGATAAATAGACAAAGAAAGAAAGACAAAGATGAAAAATACTCTCCAAAAAATGT 2421

QY 258 GluLeuProArgLeuSerGluThrSerIleLys-----AspArgMetAla 272
DB 2422 AAACITCGGCGCTGTGCAAAACCACCATTTGACACAAATCCATCTCTCGAAATGGTATCC 2481

QY 273 LysTyrGlnAlaAlaValSerLysGlnSerSer-----284
DB 2482 AAATGGATCTCACTGATGCCAAAAAACTCTGTACTGTCTATTAAGTCCATAGAAATT 2541

QY 285 ThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIleHisLysMetGlu 304
DB 2542 ACTTCGATCCTTAATGGACTTCAAGCTTCTGAAGTTCGTGTAAGACACAGTGCAGGAA 2601

QY 305 GlnLysGluAsnValProGlyProGluValCysIleThrHisGlnGluGlyLys 324
DB 2602 GATGAGAGAGGTCTCAACACATGATGATTAATATGCAAGAGAAATCTAAGATTGATCAT 2661

QY 325 IleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArg 344
DB 2662 TTGACCAACAACAGAAATGATCTTATT-----TCAAGGAGGAGAAACAGAAC 2706

QY 345 AspSerGlnValLysSerGlu-----ValGlnGlnProVal 356
DB 2707 AGTTCATCTTCTTAGAAGAAACAAAGTTTCATGCAGATTTTGGTAAATATCCAAACACG 2766

QY 357 HisProLysPro-----LeuSerProAspSer-----365
DB 2767 TCAAAATCTCCAGAAAGATTAAAGAAAGATATAGAAGTATTATCCGAAGATCTATTAT 2826

QY 366 -----ArgAlaSerSerLeuSerGluSer 373
DB 2827 GAAGAAGATGAAGTCACAAAAAGAGAAAGGATGTCAAGAGGACACAAACAGATAAATCT 2886

QY 374 SerProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCys-----Val 391
DB 2887 TCAAAACCA-----CAAATAAACCTGTGTAAAGAGGATTTGCAATACAGAA 2934

QY 392 GluCysGlnLysThrValTyrPro-----MetGluArgLeuLeuAlaAsnGlnVal 409
DB 2935 GAGTGTCTAAAACTGGATCACCTGGCAAAAGGAGAGAGAGGAGGAGGAGGAGGAGGAG 2991

QY 410 PheHisIleSerCysPheArgCysSerTyrCys-----420
DB 2992 -----TCATTTGTCATGGAACACAGTAGCAACAGCTCTTCA 3027

QY 420 -----420

DB 3028 GATGAAGATGAAGAGAAACAAAGAGATGACACCACTAAGAAATAACAATGGTTG 3087

QY 421 ---AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLys 439
DB 3088 GAGGAAAAAGAAAAATCTCTACGGACAACTGGTTCTTATTACGGA-----3132

QY 440 ProHisPheAsnGlnLeu-----PheLysSerLysGlyAsnTyrAspGluGly 455
DB 3133 -----TTTTCAGAGGTGGCAGAAAAAGGATTAACCTTTTAAATAACTCTGTATGAAGA 3186

QY 456 PheGlyHis-----ArgProHisLysAspLeuTrpAlaSer-----467
DB 3187 CTTCAAAACAGCAGCGGCCAAAGATCGAAAAAGATGCTGGTCAAGTATTCAGGGACAGTGG 3246

QY 468 ---LysAsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGlu 486
DB 3247 CCTAAAAAACGCTGAAAGAGCTTTTTTCA--GACTCTGATCTAGGCTGAGCTTCC 3303

QY 487 ThrProHisSerProGlyValGluAspAla-----ProIleAla 499
DB 3304 CCACCGCAT---CCTGCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3360

QY 500 LysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGlu 519
DB 3361 GAAGAGAGAGTGTGTCACCCAGTGTAGAACTAGAAAAAACCCACCTCCAGTCAATGTGAT 3420

QY 520 AspLysProAlaGluThrLysLysLeuArgIle-----AlaTrpPro---533
DB 3421 AGTAACCCATTGAAGAAAAACAGTAGAGGTCAATGACAGAAAAACAGAAATTTCCAAGT 3480

QY 534 -----ProProThr-----GluLeuGlySerSer---541
DB 3481 AGTGGCAGTAATTCAGTGTCTAAATACCCCTCTCTACTACACCTGAATCGGCTTCATCAGTC 3540

QY 542 -----GlySerAlaLeuGluGlyIleLysMetSerLysProLysTrpPro 557

```
Db 3541 ACTGTAACAGAGGCGGCGGAGCTCTCTGTAACAGATATCAGAACCACCTGGCTCCA 3600
Qy 558 ProGluAspGluIleSer-----LysProGluValProGluAspValAspLeuAsp--- 574
Db 3601 AACCAAGAAGAGGTTTCAAGATATCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 3660
Qy 575 -----LeuLysLysLeuArgSerSerSerSerSerSerSerSerSerSerSerSer 590
Db 3661 GTTGTGGGGAGCTCCAGACCTCCAGTCTGAAGGGAAT-----AGTCGCGCA 3708
Qy 591 PheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerPro 610
Db 3709 GCAGGTTTGTGCGGAGTGTGAGCTCAAGCAGTATGATCAGCCAGACCAACATCCT 3768
Qy 611 ProIleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGlyArgVal 630
Db 3769 -----GAAAAGCCTGTACAGGTGAGAAAGAGTGAAGATGCTCAGGAGGAGGAAGT 3822
Qy 631 AlaGluArgLysGlnValGluAsnAlaLysAla-----SerLysLysAsnGly 646
Db 3823 TCATCAAAAAGAGCAAAAGAGCCATTAAGCAGTGGTAAACAAACAAAAGAGGGA 3882
Qy 647 AsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSer 666
Db 3883 AAAGGCACAAATAGTAGTATGATGAGCAAGACTTTTCAGCTGTGTGAAGTATTAAGT 3942
Qy 667 Lys-----GluGlyHisSerLeuGlu-----MetGlu 675
Db 3943 CAGCCAGTCAAAATCAGTTTCCAGTGAAGTCTCATAGTACCAAAATCTCCGCGAAG 4002
Qy 676 AsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLys 695
Db 4003 ACGCAGTCTCCAGGAAATGTGGAAGATGGTGAAGTATGATAGGATCTGAT-----CTCAAG 4056
Qy 696 GlnGlnSerProGlnGluProLysSerLeuAsnTrpSer-----SerPheValAsp 712
Db 4057 GAACCCAGTAAATCGATTACCCAAAGTTTACAAATGGAGTTTTCAGATGTCGACCTGAA 4116
Qy 713 Asn---ThrPheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeu 731
Db 4117 AATATGACAGTGGCGAAGCATCACAATTTTCAAGAAAAATTTCAAGAAATCAGAAAA 4176
Qy 732 TrpGluGlyGluValValLysGluLeu---SerValGluGluGlnIleLysArg 748
Db 4177 CATTATCTGTCATTAATAATCGAAGTAGCTTCCATTGTCGAGGAGGAAGCGT 4230

RESULT 10
US-09-282-146-1
; Sequence 1, Application US/09282146A
; Patent No. 6303847
; GENERAL INFORMATION:
; APPLICANT: KAWAKURA, Akiyoshi
; APPLICANT: EBINUMA, Hiroyasu
; TITLE OF INVENTION: TRANSCRIPTION FACTOR CONTROLLING PHENYLPROPANOIC
; FILE REFERENCE: 4859-0027-0
; CURRENT FILING DATE: 1999-03-31
; EARLIER FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 988
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (100)..(702)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (127)..(282)
```

```
; OTHER INFORMATION: LIM domain
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (427)..(582)
; OTHER INFORMATION: LIM domain
US-09-282-146-1
```

Alignment Scores:

```
Pred. No.: 2,77e-11 Length: 988
Score: 212.50 Matches: 51
Percent Similarity: 46.58% Conservative: 24
Best Local Similarity: 31.68% Mismatches: 56
Query Match: 5.41% Indels: 30
DB: 4 Gaps: 4
```

US-09-890-549-4 (1-759) x US-09-282-146-1 (1-988)

```
Qy 384 AlaProAlaArgGluThrCysValGluCysGlnLysThrValTrpMetGluArgLeu 403
Db 109 GCAGGAACACACAGAAATGCGATGTCATGCAAGACTGTCTATCTGGTTGACAATTA 168
Qy 404 LeuAlaAsnGlnGlnValPheHisLysSerCysPheArgCysSerTyrcysAsnAsnLys 423
Db 169 ACTGCAGATAACAGAAATCTATCACAAGCTTGTTCAGATGCCATCCTGCAAGGCGCACT 228
Qy 424 LeuSerLeuGlyThrThrAlaSerLeuHisGlyArgIleTyrcysLysProHisPheAsn 443
Db 229 GTCAAGCTTGGCACTACAAATCCCTTTGAGGAGTCTTATCTATGTAGCACACCTTTGAT 288
Qy 444 GlnLeuPheLysSerLysGlyAsnTyrcysGluGlyPheGlyHisArgProHisLysAsp 463
Db 289 CAGCTCTTCAAAACAACTGGCAGTTTGGATAAAGCTTTGAAGGTACACCA-----339
Qy 464 LeuTrpAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsn 483
Db 340 -----AAAAATGTGAAG-----351
Qy 484 AlaArgGluThrProHisSerProGlyValGluAspAlaPro---IleAlaLysVal---501
Db 352 -----CCACAGAAACCCATTGACAGTGAAGAACACACAGGTAGCCAAAGTGACA 399
Qy 502 -----GlyValLeuAlaAlaSer-MetGluAlaLysAlaSerSerGlnGlnGln 517
Db 400 AGCATGTTTGGTGGAAACAGAGAGAAATGTTTGGCTGCAAGAAACTGTCTTACCAACA 459
Qy 517 uLysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGln 537
Db 460 GAAAGAGGTATCAGCAATGCGCAGCCATACCATAGAGCTGCTTCAATGCAAGCCACGGA 519
Qy 537 u 537
Db 520 G 520
```

RESULT 11

```
US-08-934-627B-1
; Sequence 1, Application US/08934627B
; Patent No. 6169174
; GENERAL INFORMATION:
; APPLICANT: OSAMU HASEGAWA
; APPLICANT: SATOSHI AOTSUKA
; APPLICANT: SOICHIRO TAKENISHI
; APPLICANT: HIROFUMI UCHIMIYA
; TITLE OF INVENTION: COTTON PLANT GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew, LLP
; STREET: 3424 Peachtree Road, N.E., 2400 Monarch Tower
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30326
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,627B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Roger T. Frost
REGISTRATION NUMBER: 22,176
REFERENCE/DOCKET NUMBER: 20111-0010
TELEPHONE: (404) 949-2400
TELEFAX: (404) 949-2499
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1015
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Gossypium hirsutum L.
FEATURE:
NAME/KEY: CDS
LOCATION: 134..757
US-08-934-627B-1

Alignment Scores:
Pred. No.: 1,4e-10 Length: 1015
Score: 205.50 Matches: 44
Percent Similarity: 52.00% Conservative: 21
Best Local Similarity: 35.20% Mismatches: 59
Query Match: 59 Indels: 1
DB: 3 Gaps: 1

US-09-890-549-4 (1-759) x US-08-934-627B-1 (1-1015)

QY 379 MetLysPheGlnAlaProLysGluThrCysValGluCysGlnLysThrValTyr 398
Db 134 ATGGGACATTCACAGGA---ACTCAACAAATGCAATGCGATGCAACAGACTTTAT 190
QY 399 ProMetGluArgLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSer 418
Db 191 CTGGTGGAATGAGTACGGCTGATTAATAGGGTTTTCATAGGCTGCTCGTTGCCAC 250
QY 419 TyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCys 438
Db 251 CATTCACAGGTTACCTCAAGCTTAGCAACTACAACTCATTCATTGAAGGGTGCTACTG 310
QY 439 LysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHis 458
Db 311 AGGCCACACTATGATCACTCTTCAAGAGAACTGCGACTCTTGACAAAGTTTGAAGGA 370
QY 459 ArgProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluLeuLysGluArgPro 478
Db 371 ACACCAAGGTTGTCAAACTGAAAGACAAATCGATAGTGAAGTGCCTCAAGAGTATG 430
QY 479 AlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaPro 498
Db 431 AACTCGTTGGTGGACAGAGAAATGTGGCGCTGTCTTAAGAGTCTTATCCAATT 490
QY 499 AlalysValGlyVal 503
Db 491 GAGAGGGTTACTGTG 505

RESULT 12

US-09-620-312D-134
; Sequence 134, Application US/09620312D
; Patent No. 6569662

GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-bong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundong
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 134
LENGTH: 4323
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (81)..(3920)
US-09-620-312D-134

Alignment Scores:
Pred. No.: 1,34e-08 Length: 4323
Score: 196.00 Matches: 154
Percent Similarity: 33.49% Conservative: 125
Best Local Similarity: 18.49% Mismatches: 324
Query Match: 4.99% Indels: 230
DB: 4 Gaps: 35

US-09-890-549-4 (1-759) x US-09-620-312D-134 (1-4323)

QY 13 SerLeuSerLeuArgValThrAlaLysGluLeuSerLeuValAsnLysAsnLysSerSer 32
Db 636 AACTTAATGTGCAAGCAAGCAAGCAAGCCTTGAAGAAAGCTACA---GGGAAGGATTCA 692
QY 33 AlaIleValGluIlePheSerLysTyrGlnLysAlaAlaGluThrAsnMetGluLys 52
Db 693 CCATGCTCAGCATCGCAGGACCCCTCCAGCAGATCCACA---GATTCAGTATGGAAATTC 749
QY 53 LysArgSerAsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeu 72
Db 750 TCAATTCC---ACTGAATGCTGATGATGAACCCGAAAGAAATGATCAACAGAGAGAA 806
QY 73 LysLysLys-----TTPGluAsnProGlyLeuGlyAlaGluSerHis 86
Db 807 AACCTTCGGACATAACTATCTTTTGTGATTTCTCCAGGACAGATGATGAAAAACAGT 866
QY 87 ThrAspSerLeuArgAsnSerSerThrGluIleArgHisArgAlaAspHisProAla 106
Db 867 GTAGATAGTGTCAAAATTCCTACTGTAGCCATAAATCTCGA-----CCTGTTTCA 917
QY 107 GluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGlnGlnGlnIleHisPro 126
Db 918 AGAGTTACCAATGGAACCTTCCAAT-----AAAAAAGTATTATGNA 959
QY 127 ArgSerArgLeuArgSerProGluAlaLeuValGlnGlyArgTyrProHisIleLys 146

Db 959 ----- 959
Qy 147 AspGlyGluAspLeuLysHisSerThrGluSerLysLysMet-----GluAsnCys 164
Db 960 -----CAAGACACTAATGATAAATACAGTACTAATAAGAAAGTCAAGTGGCAAGGATGT 1013
Qy 165 LeuGlyGluSerArgHisGluValGluLysSerGluLysSerGluAsnThrAspAlaSer 184
Db 1014 AGTGACCCAGTACCACAGGCAATTTTGAAGAAAGAGGAACCTAGCAATGGATGTACTGCA 1073
Qy 185 GlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMetPheGluLysGly 204
Db 1074 GCTCAGCAGACGACAAAGAGTAGTACCCCACTAATCTTACTATAAAGTCAAGGATCCCAAGGA 1133
Qy 205 Glu-----ProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSerGlyArgLys 223
Db 1134 GAGTCACCAACTCAGTAAAG-----TCTTCAGTCTCTTCAAGGCAG 1175
Qy 224 IleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGlnLeuSerSer 243
Db 1176 TCTGATGAATATGCGCAAGTTGGAC----- 1202
Qy 244 SerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuProArgLeuSer 263
Db 1203 CACAATACAACTACAGAGAAACACACCTAAGAGAAAAATG----- 1244
Qy 264 GluThrSerIleLysAspArgMetAlaLysTyrGlnAlaIleValSerLysGlnSerSer 283
Db 1245 GTCAAGCAAGTACACACAGCTTTGCTTAAGTTAATGCAAAAAATAGTGGCAATCCCTAAA 1304
Qy 284 SerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIleHisLysMet 303
Db 1305 AATCTAAATCAGTCAAAAAA-----GGTGAA-----ACTTTGAATAATAA 1346
Qy 304 GluGlnLysGluAsnValProGlyProGluValCysIleThrHisGlnGluGlyGlu 323
Db 1347 GATTCAAAACAGAAATGCTCTCGACAGGTATATCAAAAACTCAGCCTTCCTCCCAA 1406
Qy 324 LysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGluAspAspSer 343
Db 1407 AGACCTTTAAACATGAAACATCTACTGTCCAAAAAGT-----ATGTTTCATGATGTG 1460
Qy 344 ArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysProLeu----- 361
Db 1461 CGTGATAATAACAACAGGACAGTGTCTTCTGAACAGAGGCTCACAACCTCTCATTAAT 1520
Qy 362 -----SerProAspSerArgAlaSerSerLeuSerGluSerSerProLysAlaMet 379
Db 1521 CTTCATCTGAAATAGTATGTCAGAGCACTCCAGTCATCTCTCGAGGCTGACCCACAA 1580
Qy 380 Lys----- 380
Db 1581 AAGCCATTAAACGATCAAGAAAAAGAGAGTGGCGTTAGATGCCAAAAATATTTCAAAG 1640
Qy 381 -----LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLys 395
Db 1641 CTGGATAAATCAATTAACACAGCACTGGAATCAAAACAGATTTGTTTAGATAAAGTGAA 1700
Qy 396 ThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPhe 415
Db 1701 ACAAAATTTCCCAATCAAAAGAAACAGATGATGGCGATCGACCTAACATATGTTGTAT 1760
Qy 416 ArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArg 435
Db 1761 -----TCTGTGGAGT-----GATTAATGTAATTCANAA 1790
Qy 436 IleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGly 455
Db 1791 TTTTATAGCACACAGCCCTAAAAATACATGGTTTCAATCCAAAT-----GAAACTCC 1844
Qy 456 PheGlyHisArgProHisLysAspLeu-----TTPAlaSerLysAsnGluGluIle 474

Db 1845 TTGAACCTCTTAATCCAGTTTGTGATTTAGACTCAACAAGTGCAGGGCAAAATCCATTGTGATA 1904
Qy 475 LeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSer-----ProGly 492
Db 1905 TCAGATAGGAGACCAAGTAGGAGGAAAGATACAAACAAACAATCAAGTATTAAATGT 1964
Qy 493 ValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAla 512
Db 1965 GTGAAGATGTTTCACTGTGTAAATCTCTGAAGGACAAATGGTACCTTAAT----- 2015
Qy 513 SerSerGlnGlnGluLysGluAspLys-----ProAlaGluThrLysLysLeuArgIleAla 531
Db 2016 TCTGCTCAAGAACACAAAAATCGAAAGTTCTCTGTGGAAGGACTGACAAT----- 2066
Qy 532 TrpProProThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyLysMet 551
Db 2067 -----CCTAGTAAGTTG-----TCAGATGAATCTGCTATGGATGAA----- 2102
Qy 552 SerLysProLysTrpProProGluAspGluIleSerLysProGluValProGluAspVal 571
Db 2103 -----GACAAACATGCTCAGCA 2120
Qy 572 AspLeuAspLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPhe 591
Db 2121 GACTCAGATGATCTTCCAAAGTGTCTTCGGACAGCTATCAGAAAAAAATCTCCTAAA 2180
Qy 592 ThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPro 611
Db 2181 AATATGGAACATCAGAAATCTCCAGAGACCTATGAAACTCCAGAAACT-----CCA 2231
Qy 612 IleArgLysGlyTrpSerMetSerGluGlnSerGluGluSerValGlyGlyArgValAla 631
Db 2232 TTGTGCGTCACTGGATTTGAGTACT-----GGTGTCTGCAAT 2270
Qy 632 GluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThr 651
Db 2271 CAGCGAGAGAGTCTGAATCT----- 2291
Qy 652 ThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSer 671
Db 2292 -----GACACTGGCAGTGTACCACCTCTCCCGATGAC 2324
Qy 672 LeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAspAsn 691
Db 2325 ATAAAGCCAGATCTGAAGACTATGATCTGGAGGCTCTCAGGATGATGGTCAAT 2384
Qy 692 -----SerPheLeuLysGlnGlnSer 698
Db 2385 GACAGAGGTATCTTAATGTGGCACTATGCTGCCATGATTTCTTGGAGAGAGTACG 2444
Qy 699 -----ProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsn 713
Db 2445 AGTGATACCAGTACTCTTGAAGAATTAATA-----ATTATGATAGT 2486
Qy 714 ThrPheAlaGluLupPheThrThrGlnAsnGlnLysSerGlnAspVal----- 729
Db 2487 AATTTAAGATTAAGTAGTAAATAAGCAAGCAAGTAAATGATGATCTTTTCCAGTTAAT 2546
Qy 730 -----GluLeuTrpGluGlyGluValVal 737
Db 2547 TCAACGAGTGTATGATAATCTTAGGAAAAAGGCCAGAAATTTGGTCTCGATCTGCAATA 2606
Qy 738 LysGluLeuSerValGluGluGlnIleLysArgAsnArgTyrTyr----- 752
Db 2607 GTTCACTCTAGGAAAGAGAAATATCCAGAGGAGTGTCCAGTTGGTCTCAGGAAATA 2666
Qy 753 -----AspGluAspGluAspGlu 758
Db 2667 GATCAGGTATCTTCTTTCAGCAGATGAAACACAGATGAA 2705

RESULT 13
US-09-220-132-79
; Sequence 79, Application US/09220132


```

; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 5857
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-220-132-79

```

[illegible]

```
QY 564 LysProGluValProGluAspValAspLeu-----AspLeuLysLysLeu 578
Db 3082 AAAAAAGCATGAGGAGAAAGAAAGAAATTTGAGAGGAATTTGCGACCTGGAAAAAGAAA 3141
QY 579 ArgArgSerSer-----SerLeuLysGluArgSerArgProPheThr--- 592
Db 3142 ATGGAACCAAGCCACACACCACTGTGAGAGCTGAAAGCCAGGTATGAGAGAGCCACTTCT 3201
QY 593 -----ValAlaLaserPheGlnSerThrSerValLysSer 604
Db 3202 GAGACAAAAACCAAGCATGAAAGAAATCTCAGAACTCTCAGAAAGCGCTGCTGACACA 3261
QY 605 ProLysThrValSerProProIleArgLysGlyTrpSerMetSerGluGlnSerGluGlu 624
Db 3262 GAGGACAAAGTGAAGGCGCAGGAGGAGCAAGTGGCTGTGCGAGGAGCTGGAGGAG 3321
QY 625 SerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerLysLys 644
Db 3322 CTG-----AGAAAGCAAGCCAGCAAGCAAGCAAGCAAGTGTCTCAACA 3360
QY 645 AsnGlyAsnValGlyLysThrTrpGlnAsnLysGluSerLysGlyGluThr----- 662
Db 3361 GCGGAAGATGCCATGCAGATAATGGAAACAGATGACCAAGAGAGAGACTGAGACTCTGCC 3420
QY 663 -----GlyLysArgSerLysGlyGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 3421 TCCTTGGAGGACACCAAGCAAAATGCAAACTGCAAACTGCAAACTGCTGAGTAGAG 3480
QY 581 GluAsnGlyAlaAspSerAspGluAspAsn----- 691
Db 3481 GAAACAACTTGAAATGTGGAAGAGCTGAAACAAATCAAACTGAACTGCTGAGTAGAG 3540
QY 692 -----SerPheLeuLysGlnSerProGln 700
Db 3541 AATCAAAAAATGGAAGAAATTTAGAAAGAAATAGAAACCTTAAAGCAGCAGCTCAG 3600
QY 701 GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluPhe--- 719
Db 3601 AAGTCCAGCAGCTTTCAGCGTTTGAAGAGAGAAAGCTTAACTTCTGAGGAGCTGGGG 3660
QY 720 ThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGlu 739
Db 3661 AGAAGCAGGAGCAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 3720
QY 740 LeuSerValGluGluGlnLysArgAsnArgTyTrpAspGluAspGluAspGluGlu 759
Db 3721 AATCAGTTGTTAGAAATGAAAGAAAGAGAAATCAAGTTCATAAAGACGACATCAAGAG 3780

RESULT 14
US-08-452-655B-1
; Sequence 1. Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF INVENTION: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: DP2.5(APC)
US-08-452-655B-1

Alignment Scores:
Pred. No.: 6,54e-08 Length: 8532
Score: 194.00 Matches: 153
Percent Similarity: 35.44% Conservative: 127
Best Local Similarity: 19.37% Mismatches: 289
Query Match: 4.94% Indels: 221
DB: 1 Gaps: 38

US-09-890-549-4 (1-759) x US-08-452-655B-1 (1-8532)
QY 26 ValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLysLysGlnLysAlaAla 45
Db 5365 GTAAGAAAAAATGCAGAC-----TCAAAAAATAATTTAAATGCT 5403
QY 46 GluGluThr-----AsnMetGluLysLysArgSerAsnThrGluAsnLeuSerGln 62
Db 5404 GAGAGAGTTTCTCAGACAAACAAAGATTCAAGAAACAGAAATTTGAAAAATAATTTCAAG 5463
QY 63 HisPheArg----- 65
Db 5464 GACTTCATGATGAAGTCCCAAAATAATAGATAGAGTACAGAGAAAGTTTGTCTTTTGTAT 5523
QY 66 -----LysGlyThrLeuThrValLeuLysLysLysLysTyrGlu 77
Db 5524 TCACCTCATCTACACGCTATTGAGGAACTCTTACTGTTTTTCACGAAATGATTCT 5583
QY 78 AsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIle 97
Db 5584 TTGAGTTCTCTAGATTTTGATGATGATGATGTTGACCTTTCCAGGGAAGAGCTGATTA 5643
QY 98 ArgHisArgAlaAspHisProAlaGlu-----ValThrSerHis-----Ala 112
Db 5644 AGAAGCGAAAAGAAAATAAGAAATCAGAGGCTTAAGATTACCAAGCCACACAGAACTTAACC 5703
QY 113 AlaSerGlyAlaLysAlaAspGlnGluGluGlnIleHisProArgSerArgLeuArgSer 132
Db 5704 TCCAAACCAATCAGCTAATAAGACACAAAGTATTGAAAGCAGCCCAATAAATCAGGT 5763
```


APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA

ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/450,582

FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/452,655
FILING DATE: 25-MAY-1995
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

```

LENGTH: 8532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

```

MOLECULE TYPE: cdna
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:

CLONE: DP2.5 (APC)
1-450-582-1
ment Scores:

No.:	6,546-08	Length:	8532
..:	194.00	Matches:	153
ent Similarity:	35.4%	Conservative:	127
Local Similarity:	19.3%	Mismatches:	289
Match:	4.94%	Indels:	221
	3	Gaps:	38

-890-549-4 (1-759) x US-08-450-582-1 (1-8532)

26 ValAsnLysAsnLysSerSerAlaIleValGluLePheSerLysTyrGlnLysAlaAla 45
||| |||||
5365 GTAAGAAAAAATGCAC-----TCAAAANAATAATTAAATGCT 5403

46 GluGluThr-----AsnMetGluLysArgSerAsnThrGluAsnLeuSerGln 62
||| :::: |||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
5404 GAGACAGTTTCTCAGACAACAAGATTCAAGAACAACAGAAATTCGAATAAATTCCAAG 5463

63 HisPheArg----- 65
5464 GACTTCATGATAGCTCCCAATAATGAGATAGACTCAGAGGAAGTTTGCTTTTGAT 5523

65 -----LysGlyThrLeuThrValLeuLysLysLysTrpGlu 77
 :::|:::
5524 TCACCTCATTTACAGCCTATTGAAGCACTCCTACTGTTTTTCACGAAATGATCT 5583

QY	78	AsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIle	97
Db	5584	TTGAGTTCTCTAGATTTTGGATGATGATGTTGCACCTTTCCAGGGAAGGCTGAATTA	5643
QY	98	ArgHisArgAlaAspHisProProAlaGlu-----ValThrSerHis-----Ala	112
Db	5644	AGAAGGCAAAAGAAAATAAGGAATCAGAGCGCTAAAGTTACCAGCCACACAGAACTAAC	5703
QY	113	AlaSerGlyAlaLysAlaAspGlnGluGluHisProArgSerArgLeuArgSer	132
Db	5704	TCCAAACCAATCAGCTAATAGACACAGCTATTGCCAAGCAGCAATAAATCGAGT	5763
QY	133	ProProGluAlaLeuValGlnGlyArgTyrProHisIleLysAspGlyGluAspLeuLys	152
Db	5764	CAGCTAAACCCATACCTCAGAAACAATCCACCTTTCCCCAGTCATCCAAAGACATACCA	5823
QY	153	AspHis---SerThrGluSerLysLysMetGluAsn-----	163
Db	5824	CACAGAGGGCAGCAACTGATGAAAAGTTACAGAAATTTTGCTATTGAAAATACTCCAGTT	5883
QY	164	CysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsnThrAspAla	183
Db	5884	TGCTTTTCTATAATTCCTCTCAGTTCCTCAGTGACATTTGCTATGACCAAGAAACAAC	5940
QY	184	SerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMetPheGluLys	203
Db	5941	-----AATAAAGAAAATGAACCTATCAAGAGACACTGAGCCCCCTGACTCACAG	5988
QY	204	GlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSerGlyArgLys	223
Db	5989	GGAGAACAAGTAACCTCAA-----GCATCAGGC-----	6018
QY	224	IleSerGluAsnSerTyrSerLeuAspAspLeuGluIle-----GlyProGlyGln	240
Db	6019	TATGCTCCTAAATCATTTTCATGTTGAAGATACCCAGTTTGTTCTCAAGAAACAGTTCT	6078
QY	241	LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro	260
Db	6079	CTCAGTTCTCTAGTATTGACTCTGAAGATGAC-----	6111
QY	261	ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaValSerLys	280
Db	6112	---CTGTTGCAAGATGTATAAGCTCCGCAATGCCAATA-----AAGAAA	6153
QY	281	GlnSerSerSerThrAsnTyrThrAsnGluLeu-----LysAlaSerGlyGly---	296
Db	6154	AAGCTTTCAAGCTCAAGGGTGATATGAAAACAACATAGTCCCAAGAAATATGGGTGGCATA	6213
QY	297	-----GluIleLys-----IleHisLysMetGluGlnLysGluAsn	308
Db	6214	TTAGGTGAAGACTGACACTTGATTGAAGATATACAGAGCACCAGATTCAGAACATGGT	6273
QY	309	ValProProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsn	328
Db	6274	CTATCCCTGATTTCAGAA-----AATTTTGTGAAAGCTATTACAGGAAGGT	6321
QY	329	GluAsnSerLeu-----AlaValArgSerThrPro	338
Db	6322	GCAAATTCATAGTAGTAGTTTACATCAAGCTGCTGCTGCTGCTATCTAGACAA	6381
QY	339	AlaGluAspSerArgAspSer-----GlnValLysSerGluValGln-----Gln	354
Db	6382	GCCTGCTCTGATTCA---GATTCCATCTTCCCTCGAATCAGAAATCTCTCTGGGATCA	6438
QY	355	ProValHisProLysProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSer	374
Db	6439	CCATTTTCAT-----CTTACACTGATCAAGAAGAAAACCTTTTACAAGTAATAAA	6489
QY	375	ProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGln	394
Db	6490	GGCCACGAAATTTCTAAA-----CCAGGGGAGAAAAGTACATTGGAACTAAA	6537

Search completed: January 7, 2004, 01:47:34
Job time : 153 secs

```
QY 395 LysThrValTyrProMetGluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCys 414
Db |||
6538 AAG----- 6540
QY 415 PheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGly 434
Db |||
6540 ----- 6540
QY 435 ArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGlu 454
Db |||
6541 ----- 6541
QY 455 GlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLys---AsnGluAsnGluGlu 473
Db |||
6568 GGA---GGAAAAAAGATTATATAAGTTTGATTACTGGAAAAAGTTCGATCTAATTCAGAA 6624
QY 474 Ile-----LeuGluArgProIleGln-----LeuAlaAsnAlaArg 485
Db |||
6625 ATTTCCAGGCCAAATGAAACAGCCCTTCAAGCAAAACATGCCTTCAATCTCTCGAGGCAGG 6684
QY 486 GluThrProHisSerProGlyValGluAsp-----AlaProIleAlaLys 500
Db |||
6685 ACAATGATTCATATTTCCAGGAGTTGGAATAGTCTCTCAAGTACAAAGTCTGTTCTTAAA 6744
QY 501 ValGlyValLeuAlaAlaAspMetGluAlaLysAlaSerSerGlnGlnGlu----- 517
Db |||
6745 AAAGGCCACCCCTTAAGACTCCAGCCTCCAAAAGCCCTAGTAGAGGTCAAACAGCCACC 6804
QY 518 -----LysGluAspLysProAlaGluThrLysLysLeuArgIleAlaTyrProPro 534
Db |||
6805 ACTTCTCTAGAGGAGCCAGCCATCTGTGAATCAGAATTAAGCCCTGTTCAGGCAG 6864
QY 535 ProThrGluLeuGlySerSerGlySerAla---LeuGluGluGlyIleLysMetSerLys 553
Db |||
6865 ACATCCCAATAGGTGGTCAAGTAAAGCACCTTCTAGATCAGGATCTAGAGATTCGACC 6924
QY 554 ProLysTyrProProGluAspGluIleSerLysProGluValProGluAspValAspLeu 573
Db |||
6925 CCTTCAAGACCTGCCAGCAACCATTAAGTAGACCT----- 6960
QY 574 AspLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSer-----ArgProPhe 591
Db |||
6961 ---ATACAGTCTCTCGCCGAACTCAATTTCCCTCTGTAGAAATGAATAAGTCTCTCT 7017
QY 592 ThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPro 611
Db |||
7018 AACAAATTTATCTCAACTTCCAAGGACATCA-----TCCCTAGTACTGCTTCA----- 7065
QY 612 IleArgLysGlyTyrPsrMetSerGluGlnSerGluGluSerValGlyGlyArgValAla 631
Db |||
7066 ---ACTAAGTCTCAGGTTCTGGAAAAATGTATATATATATATATATATATATATATATAT 7122
QY 632 GluArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThr 651
Db |||
7123 CAACAGAACCTTACCAACAAACAGGTTTATCCAGAAATCCAGTAGTATTCCAAAGAAGT 7182
QY 652 ThrTrpGlnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSer 671
Db |||
7183 -----GAGTCTGCCTCCAAAGGACTAAATCAGATGAATATGATGATGAGCCAAAT 7233
QY 672 LeuGluMetGluAsnGluAsnLeu-----ValGluAsnGlyAlaAspSerAspGlu 688
Db |||
7234 AAAAAGGTAGAACTTTCTAGAAATGCTTCAACTAAATCAAGTGAAGTGAATCTGATAGA 7293
QY 689 AspAspAsnSerPheLeuLysGlnGlnSer-----ProGln-GluPr 702
Db |||
7294 TCAGAAAGACCTGATTATGATACCCAGTCACTTTTCATCAAGAGAGCTCCAAAGCCCAACC 7353
QY 702 cLysSerLeuAsnTrpSerSerPheVal 711
Db |||
7354 TTAAGAGAAATTTGGAGGAATCTGCTT 7381
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 6, 2004, 22:29:06 ; Search time 606 Seconds
(without alignments)
4335.162 Million cell updates/sec

Title: US-09-890-549-4
Perfect score: 3927
Sequence: 1 MESSPFNRQWTSLSLRVTA.....LSVVEQIKRNYDEDEDEE 759

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09890549/runat_06012004_094753_19859/app_query.fasta_1.903
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09890549@cgn_1_1_232@runat_06012004_094753_19859
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
No.							

RESULT 1
US-10-117-722-45
; Sequence 45, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/10/117.722
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 45

ALIGNMENTS

1	3927	100.0	2905	13	US-10-117-722-45	Sequence 45, Appl
2	3927	100.0	2905	15	US-10-037-270-45	Sequence 45, Appl
3	3872.5	98.6	3650	11	US-09-783-732-3	Sequence 3, Appl
4	3603.5	91.8	3543	11	US-09-783-732-1	Sequence 1, Appl
5	2535	64.6	2749	9	US-09-822-849A-399	Sequence 399, App
6	1254	31.9	732	9	US-09-925-297-152	Sequence 152, App
7	841.5	21.4	688	15	US-10-198-846-10895	Sequence 10895, A
8	828	21.1	585	15	US-10-060-036-4338	Sequence 4338, Ap
9	608.5	15.5	2710	13	US-10-094-749-1019	Sequence 1019, Ap
10	608	15.5	3236	10	US-09-833-381-1727	Sequence 1727, Ap
11	596.5	15.2	3825	11	US-09-909-567B-2	Sequence 2, Appl
12	546.5	13.9	2955	12	US-10-161-927-73	Sequence 73, Appl
13	534.5	13.6	2379	9	US-09-880-192-13	Sequence 13, Appl
14	534.5	13.6	2379	13	US-10-427-348-13	Sequence 13, Appl
15	385	9.8	1298	13	US-10-440-366-1	Sequence 1, Appl
16	383	9.8	3178	12	US-10-104-047-989	Sequence 989, App
17	381	9.7	705	10	US-09-833-381-1735	Sequence 1735, Ap
18	370.5	9.4	1299	10	US-09-833-381-1733	Sequence 1733, Ap
19	367	9.3	1081	9	US-09-789-919-5	Sequence 5, Appl
20	363	9.2	641	10	US-09-833-381-1730	Sequence 1730, Ap
21	331.5	8.4	397	9	US-09-789-919-41	Sequence 41, Appl
22	327	8.3	523	10	US-09-833-381-1736	Sequence 1736, Ap
23	290	7.4	228	10	US-09-783-580-176	Sequence 176, App
24	277	7.1	494	11	US-09-918-995-14789	Sequence 14789, A
25	273.5	7.0	743	9	US-09-789-919-6	Sequence 6, Appl
26	245	6.2	8977	12	US-10-359-012-5	Sequence 5, Appl
27	240.5	6.1	6008	12	US-10-359-012-3	Sequence 3, Appl
28	235	6.0	3290	15	US-10-084-817-286	Sequence 286, App
29	231.5	5.9	3609	10	US-09-917-800A-1715	Sequence 1715, Ap
30	226	5.8	4226	13	US-10-117-722-480	Sequence 480, App
31	226	5.8	4226	15	US-10-037-270-480	Sequence 480, App
32	224.5	5.7	861	15	US-10-278-173-83	Sequence 83, Appl
33	223	5.7	6002	13	US-10-126-704-4	Sequence 4, Appl
34	223	5.7	6002	15	US-10-071-179-4	Sequence 4, Appl
35	222	5.7	4205	14	US-10-098-841-330	Sequence 330, App
36	221	5.6	14172	12	US-10-359-012-7	Sequence 7, Appl
37	216	5.5	693	9	US-09-770-149-291	Sequence 291, App
38	215	5.5	760	15	US-10-295-403-115	Sequence 115, App
39	214	5.4	5763	14	US-10-002-600-79	Sequence 79, Appl
40	213.5	5.4	396	10	US-09-878-574-3019	Sequence 3019, Ap
41	213.5	5.4	418	10	US-09-878-574-30	Sequence 30, Appl
42	212.5	5.4	9009	12	US-09-928-412-1	Sequence 1, Appl
43	212.5	5.4	9009	12	US-10-359-012-9	Sequence 9, Appl
44	212	5.4	5659	14	US-10-002-600-80	Sequence 80, Appl
45	211	5.4	5373	12	US-10-369-493-25273	Sequence 25273, A

; LENGTH: 2905		
; TYPE: DNA		
; ORGANISM: Homo sapiens		
; FEATURE:		
; NAME/KEY: CDS		
; LOCATION: (173)..(2452)		
US-10-117-722-45		
Alignment Scores:		
Pred. No.:	0	Length: 2905
Score:	3927.00	Matches: 759
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	13	Gaps: 0
US-09-890-549-4 (1-759) x US-10-117-722-45 (1-2905)		
QY	1	MetGluSerSerProPheAsnArgGlnTTPThrSerLeuSerLeuArgValThrAla 20
Db	173	ATGGAAATCATCTCCATTTAATAGACGGCNAATGACCTCCTCATATCATTTAGGGGTAAACAGCC 232
QY	21	LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db	233	AAAGAACTTTCTTGTCAACAAGAACCAAGTCATCGGCTATTGTGGAATATTCTCCAAAG 292
QY	41	TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db	293	TACCAGAAAGCAGCTCAAGAAACAAACATGGAGAAAGAGAGAAATACACCGAAAAATCTC 352
QY	61	SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db	353	TCCAGACCTTTAGAAAGGGGACCTTGACTGTGTTAAAGAAAGAGAGTGGGAAACCCAGGG 412
QY	81	LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db	413	CTGGGAGCAGAGTCTCACACAGACTCTCTACGAAACACAGCAGCAGCTAGATTAGGCACAGA 472
QY	101	AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
Db	473	GCAGACCATCTCTCTGCTGAAGTCAAGCAGCAGCTGCTTCTGGAGCCAAAGCTGACCAA 532
QY	121	GluGlnGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db	533	GAAGAAACAAATCCACCCAGAGTCTAGACTCAGGTCACTCTCTGAAGCCCTCGTTCAAGGT 592
QY	141	ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db	593	CGATATCCCAACATCAAGACGGTGAGGATCTTAAAGACCACCTCAACAGAAAGTAAAAAA 652
QY	161	MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db	653	ATGGAAATTTGTAGGAGAAATCCAGGCATCGAAGTAGAATAATCAGAAATCAGTGAANAAC 712
QY	181	ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db	713	ACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAACATGATG 772
QY	201	PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
Db	773	TTTGAAAGGTGAACCAACTCAACTAAGATTCTTCGGGGCCCAAGCCGGAAGTGCAGT 832
QY	221	GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
Db	833	GGAAAGAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAAATAGGCCCCAGGTCAG 892
QY	241	LeuSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db	893	TTGTCATCTTACATTTGACTCGAGAAAAATCAGAGTAGAGACCAAGAAATCTGGAACTTCCA 952
QY	261	ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
Db	953	CGCTCTCAGAAACCTCTTATTAAGAGTTCGAATGCGCCAAAGTACCAGGCGAGTGTGTCCAAA 1012
QY	281	GlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
Db	1013	CAAAGCAGCTCAACCAACTATACAAATGAGCTGAAGCCAGTGGTGGCGAAATCAAAATT 1072
QY	301	HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
Db	1073	CATAAAATGGAGCAAAAGGAGAAATGTCGCCCCAGGCTCTGAGGTCTGCATCACCCATCAG 1132
QY	321	GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db	1133	GAAGGGGAAAAGATTTCTGCAATAGAAATAGCTGCGAGTCCGTTCCACCCCTGCCGAA 1192
QY	341	AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360
Db	1193	GATGACTCCCGTACTCCAGGTTAAGAGTAGGTTCAACAGAGCTGTCCATCCCAAGCCA 1252
QY	361	LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMatLys 380
Db	1253	CTAAGTCAGATTCCAGAGCTCCAGTCTTCTCTGAAAGTTCTCTCCCAAAGCAATGAAG 1312
QY	381	LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
Db	1313	AAGTTTCAGGCACCTGCAAGAGAGAGCTGCGTGGAAATGTCAAGAGACAGTCTATCCAATG 1372
QY	401	GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
Db	1373	GAGGCTCTTGTGCCAACCAAGAGGTGTTTCACTCAGCTGCTTCCGTTGCTCTCTATTGC 1432
QY	421	AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
Db	1433	AACAACAACCTCAGTCTAGAACATATGCATCTTTACATGGAAGAATCTATTGTAGACCT 1492
QY	441	HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db	1493	CAGTTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGGCACAGACCA 1552
QY	461	HisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAlaGln 480
Db	1553	CACAAGATCTATGGCCAGCAAAAAATGAAACAGAAAGAGATTTTGAGAGACAGGCCAG 1612
QY	481	LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db	1613	CTTGCAATGCAAGGGAGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTATTGTGTAAG 1672
QY	501	ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnLysGluAsp 520
Db	1673	GTGGGTGCTCTGGCTCAAAGTATGGAAGCAAGGCTCTCTCTCAGCAGCAGAGGAAGAC 1732
QY	521	LysProAlaGluThrLysLysLeuArgIleAlaTrpProProProThrGluLeuGlySer 540
Db	1733	AGCCAGCTGAACCAAGAAAGCTGAGGATCGCTCGCCACCCCTGCTGAACTTGGAACT 1792
QY	541	SerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTrpProProGluAsp 560
Db	1793	TCAGGAAGTGCCTTGGAGGAGGGATCAAAATGTCAAAGCCCAATGGCTCTCTGAAGAC 1852
QY	561	GluIleSerLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArg 580
Db	1853	GAATCAGCAAGCCGGAAGTCTCTGAGGATGTCGATCTAGATCTGAGAGAGCTAGACGA 1912
QY	581	SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
Db	1913	TCTTCTTCACTGAAGAAAGAGCGCCCATTTCACTGTAGCAGCTTCATTTCAAAGCACC 1972
QY	601	SerValLysSerProLysThrValSerProProIleArgLysGlyTyrSerMetSerGlu 620
Db	1973	TCTGTCAAGAGCCCAAAACCTGTGCCCCACCTATCAGGAAAGGCTGGAGCATGTACAG 2032
QY	621	GlnSerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
Db	2033	CAGAGTGAAGATCTGTGGGTGGAGAGTTCGAGAAAGAAACAAAGTGGAAATATCCCAAG 2092

QY 641 AlaSerLysLysAsnGlyAsnValGlyLysThrTrpGlnAsnLysGluSerLysGly 660
 Db 2093 GCTTCTAAGAAGATGGGAATGTGGGAAACCAACCTGGCAAAACAAGATCTAAGGA 2152
 QY 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluVal 680
 Db 2153 GAGACAGGAGAGAGAGTAAGGAGGTCATAGTTGGAGATGGAGATGAGATCTTGTA 2212
 QY 681 GluAsnGlyAlaAspSerAspGluAspAspAsnSerPheLeuLysGlnInSerProGln 700
 Db 2213 GAAATGGTGAGACTCCGATGAAGATGAACAGCTTCCTCAAAACAACATCTCCAAA 2272
 QY 701 GluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAlaGluGluPheThr 720
 Db 2273 GAACCAAGTCTCTGAATGGTCAGTTTGTAGACCAACCTTTGCTGAAGATTCAC 2332
 QY 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGluValValLysGluLeu 740
 Db 2333 ACTCAGAATCAGAAATCCCAAGGATGTGAACTCTGGAGGAGAGAGTGGTCAAGAGCTC 2392
 QY 741 SerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
 Db 2393 TCTGTGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAAGAG 2449

RESULT 2

US-10-037-270-45
 ; Sequence 45, Application US/10037270
 ; Publication No. US20030104529A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Chen, Rui-hong
 ; APPLICANT: Zhao, Qing A.
 ; APPLICANT: Wehrman, Tom
 ; APPLICANT: Xue, Aidong J.
 ; APPLICANT: Yang, Yonghong
 ; APPLICANT: Wang, Jian-Rui
 ; APPLICANT: Zhou, Ping
 ; APPLICANT: Ma, Yuning
 ; APPLICANT: Wang, Dunrui
 ; APPLICANT: Wang, Zhiwei
 ; APPLICANT: Tillinghast, John
 ; APPLICANT: Drmanac, Radoje T.
 ; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
 ; FILE INVENTION: Polypeptides
 ; FILE REFERENCE: 784CIP2B
 ; CURRENT APPLICATION NUMBER: US/10/037,270
 ; CURRENT FILING DATE: 2002-01-04
 ; PRIOR APPLICATION NUMBER: 09/552,317
 ; PRIOR FILING DATE: 2000-04-25
 ; PRIOR APPLICATION NUMBER: 09/488,725
 ; PRIOR FILING DATE: 2000-01-21
 ; NUMBER OF SEQ ID NOS: 1104
 ; SOFTWARE: pt_FL_genes Version 1.0
 ; SEQ ID NO 45 -
 ; LENGTH: 2905
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (173)..(2452)
 US-10-037-270-45

Alignment Scores:
 Pred. No.: 0 Length: 2905
 Score: 3927.00 Matches: 759
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Db: 15 Gaps: 0

US-09-890-549-4 (1-759) x US-10-037-270-45 (1-2905)
 QY 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
 Db 173 ATGGAATCATCTCCATTTTAAATAGACGGAATGACCTCACTATCATTTGAGGGTAAACAGCC 232
 QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
 Db 233 AAAGAACTTTCTTGTCAACAAGAACCAAGTCATCGGCTATTGTGAAATATTCTCCAAG 292
 QY 41 TyrGlnLysAlaAlaGluLysThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
 Db 293 TACCAGAAAGCAGCTGAAGAAACAACATGGAGAGAGAGAGTAACACCCGAAATCTC 352
 QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysLysLysLysLys 80
 Db 353 TCCAGCACTTTAGAAAGGGGACCTGACTGTGTTAAAGAGAGAGTGGAGAACCCAGGG 412
 QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
 Db 413 CTGGAGCAGAGCTCTCACAGACTCTCTACGGAACAGCAGCAGCTGAGATTAGGCACAGA 472
 QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
 Db 473 GCAGACCATCTCTCTGCTGAGTGAAGCCACGCTGCTTCTGGAGCCAAAGCTGACCAA 532
 QY 121 GluGluGlnIleHisProArgSerArgLeuArgSerProGluAlaLeuValGlnGly 140
 Db 533 GAAGAACAAATCCACCCAGATCTAGACTCAGGTCAGCTCCTCTGAAGCCCTGCTTCTCAGGGT 592
 QY 141 ArgTyrProHisIleLysAspGlyGluAspLysAspHisSerThrGluSerLysLys 160
 Db 593 CGATATCCCAACATCAAGGACGCTGAGGATCTTAAAGACCACTCAACAGAAAGTAAAAA 652
 QY 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
 Db 653 ATGGAAATTTGCTAGGAGATCCAGGATCAAGTAGAAAAATCAGAAATCAGTGAAGAAC 712
 QY 181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
 Db 713 ACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCGCTGAACAGGCTTAAAGATGATG 772
 QY 201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
 Db 773 TTTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCGGGCCCAAGGAGTGAAGTCAAGT 832
 QY 221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspLeuGluIleGlyProGlyGln 240
 Db 833 GGAAGGAAGATCTCTGAAAACAGCTATTCTAGATGACCTGGAAATAGGCCCCAGGTGAG 892
 QY 241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuLeuPro 260
 Db 893 TTGTCATCTTCTACATTTGACTCGGAGAAAAATGAGAGTAGACGAGAAATCTGGAACCTTCCA 952
 QY 261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaAlaValSerLys 280
 Db 953 CGCTCTCAGAAACCTCTATAAGAGATGCAATGGCCAAAGTACCAGGAGCTGTGTCCAAA 1012
 QY 281 GlnSerSerSerThrAsnTyrThrAsnGluLysLysAlaSerGlyGlyGluIleLysIle 300
 Db 1013 CAAGCAGCTCAACCACTATACAAATGAGCTGAAGCCAGTGGTGGGAAATCAAAATT 1072
 QY 301 HisLysMetGluGlnLysGluAsnValProGlyProGluValCysIleThrHisGln 320
 Db 1073 CATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCCTGAGGTCTGCATCACCCTCAG 1132
 QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
 Db 1133 GAAGGGAAAGATTTCTGCAATAGAGATAGCTGGCAGTCCGTTCCACCCCTGCCGAA 1192
 QY 341 AspAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysPro 360

Db 1193 GATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCGTGTCCATCCCAAGCCA 1252
Qy 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 1253 CTAAGTCCAGATTCACAGAGCTCCAGTCTTCTGAAAGTTCCTCCCAAGCAATGAAG 1312
Qy 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrProMet 400
Db 1313 AAGTTTCAGGCACCTGCCAGAGAGACCTCGGTGGAAATGTCAAGAACAGTCTATCAATG 1372
Qy 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCys 420
Db 1373 GAGCGTCTCTGGCCAAACAGCAGAGTGTTCACATCAGCTGCTTCGGTGTCTCTATTGC 1432
Qy 421 AsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysPro 440
Db 1433 AACAAACAAACTCAGCTAGAACATATGCATCTTTACATGGAAGAATCTATTGTAAGCCT 1492
Qy 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgPro 460
Db 1493 CACTTCATCAACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGGCACAGACCA 1552
Qy 461 HisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAlaGln 480
Db 1553 CACAAGGATCTATGGCAGAGCAAAATGAAACGAGAGATTTTGGAGAGCCAGCCAG 1612
Qy 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaLys 500
Db 1613 CTTGCAAAATGCAAGGAGACCCCTCACAGCCAGGGGTAGAAGATGCCCTATTGCTAAG 1672
Qy 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
Db 1673 GTGGGTGTCCTGGCTGCAAGTATGGAAGCCAAAGGCTCTCTCAGCAGGAGGAAGAC 1732
Qy 521 LysProAlaGluThrLysLysLeuArgIleAlaTrpProProThrGluLeuGlySer 540
Db 1733 AAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGGCCACCCCACTGAACTTGGAAAT 1792
Qy 541 SerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTrpProGluAsp 560
Db 1793 TCAGGAAGTGCTTGGAGGAGGATCAAAATGTCAAGGCCAAATGAGCTCTCAGAGAC 1852
Qy 561 GluIleSerLysProGluValProGluAspValAspLeuLysLysLeuArgArg 580
Db 1853 GAAATCAGCAAGCCGCAAGTCTCAGGATGTGATCTAGATCTGAAGAAGTAAAGCA 1912
Qy 581 SerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThr 600
Db 1913 TCTTCTTCACTGAAGGAAGAAAGCCGCCATTCATCTAGCAGCTTCATTTCAAGCAAC 1972
Qy 601 SerValLysSerProLysThrValSerProPheIleArgLysGlyTrpSerMetSerGlu 620
Db 1973 TCTGTCAAGAGCCCAAAACTGTGTCCTCCACTATCAGAAAGGCTGGAGCATGTCAAG 2032
Qy 621 GlnSerGluGluSerValGlyArgValAlaGluArgLysGlnValGluAsnAlaLys 640
Db 2033 CAGAGTGAAGAGTCTGTGGTGGAAAGATTGCAGAAAGAAACAAGTGGAAAATGCCAAG 2092
Qy 641 AlaSerLysLysAsnGlyAsnValGlyLysThrThrTyrGlnAsnLysGluSerLysGly 660
Db 2093 GCTTCTAAGAAGATGGGAATGTGGGAAAAACAACCTGGCAAAAACAAGAAATCTTAAAGGA 2152
Qy 661 GluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuVal 680
Db 2153 CAGACAGGAAGAGAGTAAAGAGGTCTAGTTGGAGATGGAGATGAGAAATCTTTGTA 2212
Qy 681 GluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGlnGlnSerProGln 700
Db 2213 GAAATGTGTGAGACTCCGATGAAGATGATAACAGCTTCCTCAACAAACATCTCCACAA 2272
Qy 701 GluProLysSerLeuAsnTrpSerPheValAspAsnThrPheAlaGluGluPheThr 720
Db 2273 GAACCCAGTCTCTGAATGTGTCAGTTTGTAGAACACACCTTTGTGTAAGATTCACCT 2332

Qy 721 ThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGlyGluValValLysGluLeu 740
Db 2333 ACTCAGAATCAGAAATCCAGGATGTGAACTCTGGAGGAGAGAGTGTCAAAGAGCTC 2392
Qy 741 SerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 2393 TCTGTGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAG 2449

RESULT 3

US-09-783-732-3
; Sequence 3, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; FILE REFERENCE: (EPLIN)
; CURRENT APPLICATION NUMBER: US/09/783,732
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3

; LENGTH: 3650
; TYPE: DNA
; ORGANISM: Homosapien
US-09-783-732-3

Alignment Scores:

Pred. No.: 0 Length: 3650
Score: 3872.50 Matches: 750
Percent Similarity: 99.08% Conservative: 3
Best Local Similarity: 98.68% Mismatches: 6
Query Match: 98.61% Indels: 1
DB: 11 Gaps: 1

US-09-890-549-4 (1-759) x US-09-783-732-3 (1-3650)

Qy 1 MetGluSerSerProPheAsnArgArgGlnTTPThrSerLeuSerLeuArgValThrAla 20
Db 102 ATGGATCATCTCCATTTAATAGCCGCAATGAGCTCACTATCATTTAGGGTAAAGCC 161
Qy 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 162 AAAGAACTTCTCTCTGTCACAGAACAAAGTCATCGCTATTGTGGAATATTTCTCAAG 221
Qy 41 TyrGlnLysAlaAlaGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 222 TACCAGAAAGCAGCTCAAGAAACAAACATGGAGAGAGTAACACCCGAAATCTC 281
Qy 61 SerGlnHisPheArgLysGlyThrIleuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 282 TCCACGACTTTAGAAAGGGGACCTTGACTGTGTTAAAGAGAGTGGGAGAACCCAGG 341
Qy 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 342 CTGGAGCAGAGTCTCACACAGACTCTTACGGACAGCAGCAGCTGAGATTAGGCACAGA 401
Qy 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
Db 402 GCAGACCATCTCTCTGCTCAAGTGACCAAGCCAGCTGCTTCTGGAGCCAAAGCTGACCA 461
Qy 121 GluGluGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGly 140
Db 462 GAAGAACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAAGCCCTCGTTCAGGT 521
Qy 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 522 CGATATCCCCACATCAAGGACCGTGGAGATCTTAAAGACCACTCAACAGAAAGTAAAAA 581

161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db ATGGAATAATTCCTAGAGAAATCCAGGATGAAAGTAAAGAAATCAGAAATCAGTGAAGAAC 641
181 ThrAspAlaSerGlyLysIleGluLysTyrAsnValProLeuAsnArgLeuLysMetMet 200
Db ACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAAACAGGCTTAAATGATG 701
201 PheGluLysGlyGluProThrGlnThrLysIleLeuArgAlaGlnSerArgSerAlaSer 220
Db TTTGAGAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCCAAAGCCGAAAGTGAAGT 761
221 GlyArgLysIleSerGluAsnSerTyrSerLeuAspAspLeuGluIleGlyProGlyGln 240
Db GGAAGAGAGATCTCTGAAACACGCTATTCTCTAGATGACCTGGAATAATAGCCCGAGTCAG 821
241 LeuSerSerSerThrPheAspSerGluLysAsnGluSerArgArgAsnLeuGluLeuPro 260
Db TTGTCATCTTCTCATTTGACTCGGAGAAATAGAGATAGACGAATCTGGAACTTCCA 881
261 ArgLeuSerGluThrSerIleLysAspArgMetAlaLysTyrGlnAlaValSerLys 280
Db CGCCTCTCAGAAACCTCTATAAAGGATCGAATGCCCAAGTACCAGGCAGCTGCGTCCAAA 941
281 GlnSerSerSerThrAsnTyrThrAsnGluLysLysAlaSerGlyGlyGluIleLysIle 300
Db CAAAGCAGCTCAACCAACTATACAAATGAGCTGAAGCCAGTGGTGGCGAAATCAAAAT 1001
301 HisLysMetGluGlnLysGluAsnValProGlyProGlyValCysIleThrHisGln 320
Db CATAAATGAGCNAAGAGGAATGTGCCCCAGGCTCTGAGGTCTGCAATCACCCTCAG 1061
321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db GAAGGGAAAGATTTCTGCAAAATAGCAATAGCTGCGACTCGTTCACCCCTGCGGAA 1121
341 AspAspSer--ArgAspSerGlnValLysSerGluValGlnGlnProValHisProLys 359
Db GATGACTCCCGAGTGACTCCAGGTTAAGAGTGAAGTTCACAGCCTGTCCATCCCAAG 1181
360 ProLeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProLysAlaMet 379
Db CCACTAAGTCCAGATCCAGAGCTCCAGCTCTTCTGAAAGTTCCTCCTCCCAAGCAATG 1241
380 LysLysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyrPro 399
Db AAGAAGTTTCAGGCACCTGCAAGAGACCTGCGTGGATGTCAGAGACAGCTTATCCA 1301
400 MetGluArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyr 419
Db ATGGAGCGTCTCTGGCCAAACAGCAGAGGTTTTCACATCAGCTGCTTCGTTGCTCCTAT 1361
420 CysAsnAsnLysLysLeuSerGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLys 439
Db TGCAACAACAACCTCAGCTAGGAAACATATGCACTTTTACATGGAAGATCTATTGTAAG 1421
440 ProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluClyPheGlyHisArg 459
Db CCTCATTCTCAATCACTCTTTAAATCTAAGGGCAACTATGATGAAGGCTTTGGGCACAGA 1481
460 ProHisLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIleLeuGluArgProAla 479
Db CCACAAAGGATCTATGGGCAACCAAAATGAAACGAGAGATTTTGAGAGAGACAGCC 1541
480 GlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAla 499
Db CAGCTTGCAAATGCAAGGAGAGACCCCTCAGCCAGGGGTAGAAATGCCCTATTGCT 1601
500 LysValGlyValLeuAlaValSerMetGluAlaLysAlaSerSerGlnGlnLysGlu 519
Db AAGTGGGTGCTCTGGCTGCAAGTATGGAAGCGCAGGGCTCTCTCAGCAGGAGAGAA 1661

520 AspLysProAlaGluThrLysLysLeuArgIleAlaTyrProProProThrGluLeuGly 539
Db GACAAAGCCAGCTGAAACCAAGAGCTGAGATCGCTGCGCCACCCCTGAACTTGA 1721
540 SerSerGlySerAlaLeuGluGlyIleLysMetSerLysProLysTrpProGlu 559
Db AGTTTCAAGAGTGCCTTGGAGGAAGGATCAAAATGTCAAAGCCCAATGGCTCTCTGAA 1781
560 AspGluIleSerLysProGluValProGluAspValAspLeuAspLysLysLeuArg 579
Db GACGAAATCAGCAAGCCGAGTCTCTGAGGATGTCGATCTAGATCTGAAGAGCTAAGA 1841
580 ArgSerSerSerLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSer 599
Db CGATCTCTCTACTGAAGGAAAGACCCGCCATTCCTGTAGCAGCTTCAATTCAAAGC 1901
600 ThrSerValLysSerProLysThrValSerProProIleArgLysGlyTyrSerMetSer 619
Db ACTCTGTCTCAAGAGCCCAAAACCTGTGTCCCACTATCAGAAAGGCTGGAGCATGTCA 1961
620 GluGlnSerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAla 639
Db GAGCAGATGAAGATCTGTGGTGGAGAGTTGAGAAAGGAAACAAAGTGGAAATGCC 2021
640 LysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTyrGlnAsnLysGluSerLys 659
Db AAGGCTTCTAAGCAATGGGAATCTGGCAAAACCAACCTGCGCAAAACAAAGATTTAA 2081
660 GlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluLeu 679
Db GGAGAGACGGGAGAGAGAGTAAGGAAGTTCATAGTTGGAGATCGAGATGAGAAATTT 2141
680 ValGluAsnGlyValaAspSerAspGluAspAspAsnSerPheValAspAsnThrPheAlaGluPhe 699
Db GPAGAAATGGTGGAGATCTCCGATGAAGATGAATACAGCTTCTCCTCAAAACAAATTTCCA 2201
700 GlnGluProLysSerLeuAsnTyrSerPheValAspAsnThrPheAlaGluGluPhe 719
Db CAAGAACCCAGTTTGTGAATGTCGAGTTTGTAGAACACACCTTCTGCTGAAGATTC 2261
720 ThrThrGlnAsnGlnLysSerGlnAspValGluLeuTyrGluGlyValValLysGlu 739
Db ACTACTCAGATCAGAAATCCAGGATCGGAATCTTGGAGGGAGAGAGTGGTCAAGAG 2321
740 LeuSerValGluGluGlnIleLysArgAsnArgTyrTyrAspGluAspGluGlu 759
Db CTCTCTGTGGAGAACAGATAAGAGAAATCGGTATTATGATGAGATCAGATCAGAG 2381

RESULT 4
US-09-783-732-1
; Sequence 1, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; TITLE OF INVENTION: (EPLIN)
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09/783,732
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3543
; TYPE: DNA
; ORGANISM: Homosapien
US-09-783-732-1
Alignment Scores:
Pred. No.: 0 Length: 3543

Score:	3603.50	Matches:	695
Percent Similarity:	99.04%	Conservative:	3
Best Local Similarity:	98.58%	Mismatches:	6
Query Match:	91.76%	Indels:	1
DB:	11	Gaps:	1
US-09-890-549-4 (1-759) x US-09-783-732-1 (1-3543)			
QY	56	AsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLys	75
DB	159	AAACCGAAATCTCTCCAGACATTTAGAAAGGGACCCCTGACTGTGTAAAGAAGAAG	218
QY	76	TypGluAsnProGlyLeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThr	95
DB	219	TGGGAGAACCCAGGGCTGGGACACAGAGTCTCACAGAGCTCTCTACGGAACAGACAGCACT	278
QY	96	GluileArgHisArgAlaAspHisProProAlaGluValThrSerHisAlaAlaSerGly	115
DB	279	GAGATTAGGCACAGACGACCATCTCTCTGCTGAGTGACAGGCCACGCTCTCTCGA	338
QY	116	AlaLysAlaAspGlnGluGluGlnHisProArgSerArgLeuArgSerProProGlu	135
DB	339	GCCAAAGCTGACCAAGAAACAAATCCACCCAGATCTAGACTCAGGTCACCTCCTGAA	398
QY	136	AlaLeuValGlnGlyArgTyrProHisLysLysAspGlyGluAspLeuLysAspHisSer	155
DB	399	GCCTCTGTTACGGTTCGATATCCCCACATCAAGGACGGTGAGGATCTTTAAAGACCACTCA	458
QY	156	ThrGluSerLysLysMetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSer	175
DB	459	ACAGAAAGTAAATAATGGAATTTCTAGAGAGATCCAGGCATGAGTAGAATAATCA	518
QY	176	GluileSerGluAsnThrAspAlaSerGlyLysLysLysLysLysLysLysLysLysLys	195
DB	519	GAATCAGTGAATAACACACATGCTTCGGCAAAATAGAGAAATATAATGTTCCGCTGAC	578
QY	196	ArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysLysLysLysLysLys	215
DB	579	AGCTTTAGATGATGTTTGAGAAAGGTGAACCACTCAAACTTAAGATTCTCCGGGCCAA	638
QY	216	SerArgSerAlaSerGlyArgLysLysSerGluAsnSerTyrSerLeuAspAspLeuGlu	235
DB	639	AGCCGAAGTGCAAGTGGAGGAGATCTCTGMAAACACAGTATCTCTAGATGACCTGGAA	698
QY	236	IleGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSerArgArg	255
DB	699	ATAGGCCACAGTGTGTCATCTTCTACATTTGACTCGAGAAATATGAGAGTAGACGA	758
QY	256	AsnLeuGluLeuProArgLeuSerGluThrSerLysLysAspArgMetAlaLysTyrGln	275
DB	759	AAATCTGGAACTTCCACGCTCTCAGAAACCTCTATAAAGGATCGAAATGGCCCAAGTACCAG	818
QY	276	AlaAlaValSerLysGlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGly	295
DB	819	GCAGTGTGCTCCAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT	878
QY	296	GlyGluileLysLysHisLysMetGluGlnLysGluAsnValProProGlyProGluVal	315
DB	879	GGCGAAATCAAAATTCATAAATGAGCAAAAGGAGATGTGCCCCAGGTCCTGAGGTC	938
QY	316	CysIleThrHisGlnGluGlyGluLysLysSerAlaAsnGluAsnSerLeuAlaValArg	335
DB	939	TGCATACCCCATCAGAAAGGGGAAAGATTTCTGCAATGAGATAGCTTGGCAGTCCGT	998
QY	336	SerThrProAlaGluAspAspSer---ArgAspSerGlnValLysSerGluValGlnGln	354
DB	999	TCCACCTCGCCGAGATGATCTCCCGAGGTGACTCCCGAGGTAAAGAGTGAAGTTCAACAG	1058
QY	355	ProValHisProLysProLysProAspSerArgAlaSerSerLeuSerGluSerSer	374
DB	1059	CCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCCTCCAGTCTTTCTGAAAGTCT	1118
QY	375	ProProLysAlaMetLysLysPheGlnAlaProAlaArgGluThrCysValGluCysGln	394

DB	1119	CCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGAATGTCTAG	1178
QY	395	LysThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPheHisLysSerCys	414
DB	1179	AAGACAGTCTATCCAATGGAGCGTCTCTTGGCCCAACAGCAGGGTGTTCACATCAGGTGC	1238
QY	415	PheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGly	434
DB	1239	TTCCGTTGCTCTTATTCGAACAACAACACTCAGTCTAGGAACATATGTCATCTTTACATGGA	1298
QY	435	ArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGlu	454
DB	1299	AGAATCTATTGTAAAGCCTCACTTCAATCAACTCTTTAAATCTAAGGCAACTATGATGAA	1358
QY	455	GlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLysAsnGluAsnGluLeu	474
DB	1359	GGCTTGGGCACAGACCACCAAGGATCTATGGCAAGCAAAATGAAACGAAGAGATT	1418
QY	475	LeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGlu	494
DB	1419	TTGAGAGACACCGCCAGCTTGCAATGCAAGGAGAGCCCTCACAGCCAGGGGTAGAA	1478
QY	495	AspAlaProIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSer	514
DB	1479	AATGCCCTTATTGCTAAGGTGGGTCTCTGGCTGCAAGTATGGAAGCCAGGGCTCTCTCT	1538
QY	515	GlnGlnLysGluAspLysProAlaGluThrLysLysLeuArgLysLysLysLysLysLys	534
DB	1539	CAGCAGAGAGAGAAAGCAAGCAGCTGAAACCAAGAGCTGAGGATCGCTCGGCCACCC	1598
QY	535	ProThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyLysLysMetSerLysPro	554
DB	1599	CCACTGMACTTGGAGATTGAGGAGTCTTGAGAGAGGATCAAAATGTCAAAGCCC	1658
QY	555	LysTrpProGluAspGluLysLysLysProGluValProGluAspValAspLeuAsp	574
DB	1659	AAATGGCTCTCTCAAGACGAAATCAGCAAGCCGAAAGTCTCTGAGGATGTCTAGAT	1718
QY	575	LeuLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPheThrValAla	594
DB	1719	CTCAAGAAGCTAAGACGATCTTCTTCAAGAGAAAGAGCCGCCCTTCACTGTAGCA	1778
QY	595	AlaSerPheGlnSerThrSerValLysSerProLysThrValSerProIleArgLys	614
DB	1779	GCCTTCAATTTCAAAGCCTCTGTCAAGAGCCCAAAACTGTGTCCCTATCAGGAA	1838
QY	615	GlyTrpSerMetSerGluGlnSerGluSerValGlyGlyArgValAlaGluArgLys	634
DB	1839	GGCTGGAGCATGTTCAGCAGCAATGAGGAATCTGTGGGTGGAAGAGTTGCAAGAAAGAAA	1898
QY	635	GlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrThrTrpGln	654
DB	1899	CAAGTGGAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTGGGAAAAACAACCTGGCAA	1958
QY	655	AsnLysGluSerLysGlyThrGlyLysArgSerLysGlyGlyHisSerLeuGluMet	674
DB	1959	ACAAAGAAATTTAAAGAGAGAGACAGGAAAGAGTAAGAAAGTTCATAGTTTGGAGATG	2018
QY	675	GluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeu	694
DB	2019	GAGATGAGAAATTTGTAGAAATGGTCAGACTCCGATGAAGATGATAAACAGCTTCTCTC	2078
QY	695	LysGlnGlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThr	714
DB	2079	AAACAACAATTTCCACAAGACCCCAAGTTTTTGAATTTGGTCCGAGTTTGTACACACACC	2138
QY	715	PheAlaGluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly	734
DB	2139	TTTGTGAAGAAATTCACACTACTCAGAAATCCAGGATGTGGAATCTTTGGGAGGGA	2198
QY	735	GluValValLysGluLeuSerValGluGlnLysArgAsnArgTyrTyrAspGlu	754

Db 2199 GAAGTGTCAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAG 2258
QY 755 AspGluAspGluGlu 759
Db 2259 GATGAGGATGAAGAG 2273
RESULT 5
US-09-822-849A-399
; Sequence 399, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fichtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; PRIORITY FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 399
; LENGTH: 2749
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-399
Alignment Scores:
Pred. No.: 7,78e-223 Length: 2749
Score: 2535.00 Matches: 490
Percent Similarity: 99.80% Conservative: 0
Best Local Similarity: 99.80% Mismatches: 1
Query Match: 64.55% Indels: 1
DB: 9 Gaps: 0
US-09-890-549-4 (1-759) x US-09-822-849A-399 (1-2749)
QY 269 AspArgMetAlaLysTyrGlnAlaValSerLysGlnSerSerThrAsnTyrThr 288
Db 1 GATCGAATGGCCAAAGTACCAGGAGCTGTGTCCAAACAAAGAGCTCAACCACTATACA 60
QY 289 AsnGluLeuLysAlaSerGlyGluLulleLysIleHisLysMetGluGlnLysGluAsn 308
Db 61 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCA.AAGGAGAA 119
QY 309 ValProProGlyProGluValCysIleThrHisGlnGluGlyGluLysIleSerIleAsn 328
Db 120 GTGCCCCAGGTCTCTGAGTCTGATCAACCATCAGGAAGGGAAGATTTCTGCAAA 179
QY 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspSerArgAspSerGlnVal 348
Db 180 GAGATAGCTGGCAGTCTGTTCCACCCCTGCGGAAGATGACTCCCGTACTCCAGGTT 239
QY 349 LysSerGluValGlnGlnProValHisProLysProLysProLysProLysProLysProLys 368
Db 240 AAGAGTGAAGTTCAACAGCCTGTCATCCCAAGCCACTAAGTCAGATTTCCAGACCTCC 299
QY 369 SerLeuSerGluSerProProLysAlaMetLysLysPheGlnAlaProAlaArgGlu 388
Db 300 AGTCTTTCTGAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCCAAGAG 359
QY 389 ThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGln 408
Db 360 ACCTGGTGGAAATGTCAGAAACAGCTCTATCAATGAGCGCTCTCTTGGCCCAACAGAG 419
QY 409 ValPheHisIleSerPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGlyThr 428

Db 420 GTGTTTTCACATCAGCTGCTTCGTTGCTCCATTGCAACAACAACTCAGTCTAGAAACA 479
QY 429 TyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSer 448
Db 480 TATGCATCTTTACATGGAAGATCTATTGTAAAGCTCTCAATCAATCTTTAAATCT 539
QY 449 LysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLys 468
Db 540 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACACACAAAGGATCTATGGCAGACAA 599
QY 469 AsnGluAsnGluGluLulleLysGluArgProAlaGlnLeuAlaAsnAlaArgGluThrPro 488
Db 600 AATGAAAAACGAAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGAGACCCCT 659
QY 489 HisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAlaSerMet 508
Db 660 CACAGCCAGGGGTAGAAGATGCCCTTATTCTAAGTGGGTCTCTGCTGCAAGTATG 719
QY 509 GluAlaLysAlaSerSerGlnGlnGluLysGluAspLysProAlaGluThrLysLysLeu 528
Db 720 GAAGCCAAAGGCTCTCTCTCAGCAGGAGAGAAAGCAAGCAGCTGAAACCAAGAGCTG 779
QY 529 ArgIleAlaTyrProProThrGluLeuGlySerSerGlySerAlaLeuGluGluGly 548
Db 780 AGGATCGCTGGCCACCCCTCACTGAAGTTCAGGAAGTCTTGGAGGAAGGG 839
QY 549 IleLysMetSerLysProLysTyrProProGluAspGluLulleSerLysProGluValPro 568
Db 840 ATCAAAATGTCAAAAGCCCAATGGCTCTGAGAGCAAGAAATCAGCAAGCCGAAAGTCT 899
QY 569 GluAspValAspLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 588
Db 900 GAGGATGTCGATCTAGATCTGAGAAAGTCAAGAGCTCTTCTTCACTGAAGGAAGAGC 959
QY 589 ArgProPheThrValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrVal 608
Db 960 CGCCCATTCATCTAGCAGCTTCTATTCAAGAGCACTCTGTCAAGACCCCAAACTGTG 1019
QY 609 SerProProLysArgLysGlyTyrSerMetSerGluGlnSerGluGluSerValGlyGly 628
Db 1020 TCCCCACCTATCAGAAAGGCTGGAGCATGTGACAGCAGAGTGAAGAGTCTGTGGGTGGA 1079
QY 629 ArgValAlaGluArgLysGlnValGluAsnAlaLysLysLysLysLysLysLysLysLys 648
Db 1080 AGAGTTCGCAAAAGGAAACAAAGTGGAAATGCCAAGGCTTCTAAGAGAAGATGGAGTGT 1139
QY 649 GlyLysThrThrTyrGlnAsnLysGluSerLysGlyGlyGlyGlyGlyGlyGlyGlyGly 668
Db 1140 GGAAAAACAACCTGGCCAAAACAAAGAAATCTAAAGGAGAGACAGGGAAGAGAAAGTAA 1199
QY 669 GlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGlu 688
Db 1200 GGTCAATGTTGAGATGAGAGATGAGAAATCTGTAGAAAATGGTGCAGACTCCGATGAA 1259
QY 689 AspAsnSerPheLeuLysGlnGlnSerProGlnGluProLysSerLeuAsnTyrSer 708
Db 1260 GATGATAACAGCTTCTCTCAAAACAAATCTCCCAAGAACCCCAAGTCTCTGAATTGTGTC 1319
QY 709 SerPheValAspAsnThrPheAlaGluGluPheThrThrClnAsnGlnLysSerGlnAsp 728
Db 1320 AGTTTTGTAGCAACACCTTTGCTGAAGAAATTCACACTCAGAAATCAGAAATCCCAAGAT 1379
QY 729 ValGluLeuTyrGluGlyGluValValLysGluLeuSerValGluGluGlnIleLysArg 748
Db 1380 GTGAACTCTGGAGGAGAGAGTGTCAAGAGCTCTTGTGGAAGAACAGATATAAGAGA 1439
QY 749 AsnArgTyrTyrAspGluAspGluAspGluGlu 759
Db 1440 AATCGGTATTATGATGAGGATGAGGATGAAGAG 1472
RESULT 6
US-09-925-297-152

```
; Sequence 152, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCI/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 152
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (729)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-152

Alignment Scores:
Pred. No.: 5,368-106 Length: 732
Score: 1254.00 Matches: 238
Percent Similarity: 98.76% Conservative: 0
Best Local Similarity: 98.76% Mismatches: 3
Query Match: 31.93% Indels: 0
Gaps: 0

US-09-890-549-4 (1-759) x US-09-925-297-152 (1-732)

QY 321 GluGlyGluLeuLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThrProAlaGlu 340
Db 2 GAAGGGGAAGAATTTCTGCAAAAGAGATAGCTGGCAGTCCGTTCCACCCCTGCCGAA 61
QY 341 AspSerArgAspSerGlnValYserGluValGlnGlnProValHisProLysPro 360
Db 62 GATGACTCCCGTACTCCAGGTTAAGAGTGAGGTTCACAGCCGTGTCATCCCAAGCCA 121
QY 361 LeuSerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLys 380
Db 122 CTAAGTCAGATTCAGAGCCTCCAGTCTTTCTGARAGTTCTCTCCCAAGCAATGAAG 181
QY 381 LysPheGlnAlaProAlaArgGluThrCysValGluCysGlnLysThrValTyProMet 400
Db 182 AAGTTTCAGGCACCTGCGAGAGACCTGCGTGAATGTGAGAGACAGTCTATCCAAATG 241
QY 401 GluArgLeuLeuAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrcys 420
Db 242 GAGCGTCTCTGGCCAAACCAGCAGGTGTTTACATCAGCTGCTTCGTTCTCTCTATTGC 301
QY 421 AsnAsnLysLeuSerLeuGlyThrTyAlaSerLeuHisGlyArgIleTyrcysLysPro 440
Db 302 AACCAAAATCACTCTAGGAACATATGCATCTTTACATGAAGAATCTATTGAGCCT 361
QY 441 HisPheAsnGlnLeuPheLysSerLysGlyAsnTyrcysGluGlyPheGlyHisArgPro 460
Db 362 CACTTCAATCACTCTTTAAATCTAAGGCAACTATGATGAAGCCTTTGGGCACAGACCA 421
QY 461 HisLysAspLeuThrAlaSerLysAsnGluGluLeuGluAlaGln 480
Db 422 CACAAGGATCTATGGCAAGCAAAAATGAACAGAGAGATTTGGAGAGACCCAG 481
QY 481 LeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProLys 500
Db 482 CTTGCAATCAAGGGAGAGCCCTCACAGCCAGGGGTAGAGATGCCCTATTGCTAAG 541
QY 501 ValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp 520
Db 542 GTGGGTCTCTKKCTGCAAGTATGGAAGCCCAAGCCCTCTCTCAGCAGGAGGAAGAC 601

; Sequence 152, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Streimann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10895
; LENGTH: 698
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1 2, 691, 692, 693, 694, 695, 696, 697, 698
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10895

Alignment Scores:
Pred. No.: 4,748-68 Length: 698
Score: 841.50 Matches: 169
Percent Similarity: 90.91% Conservative: 1
Best Local Similarity: 90.37% Mismatches: 2
Query Match: 21.43% Indels: 16
Gaps: 1

US-09-890-549-4 (1-759) x US-10-198-846-10895 (1-698)

QY 1 MetGluSerSerProPheAsnArgGlnTrpThrSerLeuSerLeuArgValThrAla 20
Db 174 ATGGAATCATCTCCATTAAATAGACGGCAATGGACCTCATTATCATTCATGAGGTAAACGCC 233
QY 21 LysGluLeuSerLeuValAsnLysAsnLysSerSerAlaIleValGluIlePheSerLys 40
Db 234 AAAGAACTTCTCTGTGTCAACAAGAACAGTCATCGGCTATTGTGAAATATTCTCCAA- 292
QY 41 TyrGlnLysAlaAlaGluGluThrAsnMetGluLysLysArgSerAsnThrGluAsnLeu 60
Db 293 -----AACACCGAAATCTC 307
QY 61 SerGlnHisPheArgLysGlyThrLeuThrValLeuLysLysLysTrpGluAsnProGly 80
Db 308 TCCAGAGACTTTAGAAAGGGACCCCTGCTGTGTGTAAAGAGAGAGTGGGAGAACCCAGGG 367
QY 81 LeuGlyAlaGluSerHisThrAspSerLeuArgAsnSerSerThrGluIleArgHisArg 100
Db 368 CTGGGAGCAGAGTCTCACAGACTCTCTACGGAACAGCAGCTGAGATTAGGCACAGA 427
QY 101 AlaAspHisProProAlaGluValThrSerHisAlaAlaSerGlyAlaLysAlaAspGln 120
```

Db 428 GCAGACCATCTCTGCTGAAGTACAAAGCCACGCTCTCTGAGCCAAAGCTGACCAA 487
Qy 121 GluGluGlnIleHisProArgSerArgLeuArgSerProGluAlaLeuValGlnGly 140
Db 488 GAAGAAACAATCCACCCAGATCTAGACTCAGGTCACTCTCTGAAAGCCCTCGTTCAAGGT 547
Qy 141 ArgTyrProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLys 160
Db 548 CGATATCCACATCAGGACGGTGAAGATCTTAAGACCACCTCAACAGAAAGTAAAAAA 607
Qy 161 MetGluAsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsn 180
Db 608 ATGAAATTTGCTAGGAGATCCAGGCATGAAGTAGAAGTAAAGATCAGATCAGTGAAC 667
Qy 181 ThrAspIleSerGlyLysIle 187
Db 668 ACAGATGCTTTCCGGGCAATA 688

RESULT 8

US-10-060-036-4338

; Sequence 4338, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yuqiu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4338
; LENGTH: 565
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 415..418, 556
; OTHER INFORMATION: n = A,T,C or G
US-10-060-036-4338

Alignment Scores:

Pred. No.:	6.17e-67	Length:	565
Score:	828.00	Matches:	162
Percent Similarity:	98.78%	Conservative:	0
Best Local Similarity:	98.78%	Mismatches:	1
Query Match:	21.08%	Indels:	1
DB:	15	Gaps:	0

US-09-890-549-4 (1-759) x US-10-060-036-4338 (1-565)

Qy 597 PheGlnSerThrSerValLysSerProLysThrValSerProIleArgLysGlyTrp 616
Db 1 TTTCAGACCATCTCTGTCAGAGCCCAAAACTGTCTCCACCTATCAGGAAAGGCTGG 60
Qy 617 SerMetSerGluIleSerGluSerValGlyArgValAlaGluArgLysGlnVal 636
Db 61 AGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGATTGAGAAAGGAACAGTG 120
Qy 637 GluAsnAlaLysAlaSerLysLysnGlyAsnValGlyLysThrThrTrpGlnAsnLys 656
Db 121 GAAATGCAAGGCTCTTAAGAAAGATGGGAATGTGGGAAAAACAACCTGCGCAAAACAA 180
Qy 657 GluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsn 676
Db 181 GAATCTTAAGGAGACAGGAGGAGAGAGTAGAGGAGGTATAGTTTGGAGATCGGAAT 240

Qy 677 GluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSerPheLeuLysGln 696
Db 241 GAGAATCTTTGTAGAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCTCTCAACAA 300
Qy 697 GlnSerProGlnGluProLysSerLeuAsnTrpSerSerPheValAspAsnThrPheAla 716
Db 301 CAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTGGAGTTTGTAGACAACACCTTTGCT 360
Qy 717 GluGluPheThrThrGlnAsnGlnLysSerGlnAspValGluLeuTrpGluGly-GluVa 736
Db 361 GAAGAATTCACCTACTACTCAGAAATCCAGGATGTGGAACTCTGGAGGAGGAGNANGT 420
Qy 736 lVallysGluLeuSerValGluGluGlnIleLysArgAsnArgTyrTrpAspGluAspG1 756
Db 421 GGTCAAGAGAGTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATCA 480
Qy 756 uAspGluGlu 759
Db 481 GGATGAAGAG 490

RESULT 9

US-10-094-749-1019
; Sequence 1019, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1019
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1019

Alignment Scores:

Pred. No.:	7.8e-46	Length:	2710
Score:	608.50	Matches:	204
Percent Similarity:	42.42%	Conservative:	115
Best Local Similarity:	27.13%	Mismatches:	279
Query Match:	15.50%	Indels:	154
DB:	13	Gaps:	27

US-09-890-549-4 (1-759) x US-10-094-749-1019 (1-2710)

Qy 12 ThrSerLeuSerLeuArgValThrAlaLysGluLeuSerLeuValAsnLysSer 31
Db 681 ACTTTTATCATAAGTACTTCTGCCCCGACGAAACAGAAAAACCTTACTAAGAACGAGCTT 740

QY 32 SerAlaIleValGluIlePheSerLysTyrGlnLysAlaAlaGluThrAsnMetGlu 51
|||
Db 741 TCTCAGTCCCTAAAAGACAGT---TATGTTGAACCCCCACCAGAGGCCCATGTCG 797
QY 52 LysLys-----ArgSerAsnThrGluAsnLeuSerGlnHisPheArgLysGly 67
|||
Db 798 CAAAAATCTGAAATTCACAGACGACAACTTCCCTCTCCACCAGAGTCGCTCTGAA 857
QY 68 ThrLeuThrValLeuLysLysLysTrpGluAsnProGlyLeuGlyAla----- 83
|||
Db 858 CAACCTGTGCAGACTCAAGACACCACTGCAAAAGTTATCCAAAGGGCCATCCCATGTCCA 917
QY 84 -----GluSerHisThrAspSerLeuArgAsnSerSerThr 95
|||
Db 918 GCAGCAACCCCGTTCCAATTTGAGAGAGAGGTCTGAAATCATCATCTCTCCCTGCAACA 977
QY 96 -----GluIleArgHisArgAlaAspHisProProAlaGluValThr--- 109
|||
Db 978 CTTCTGTCGTCAAATTAAGATAGAACTCGTGTAGGACTCTCCACCTACAAATCACAATA 1037
QY 110 -----SerHisAlaIleSerGlyAla-----LysAlaAspGlnGlu 122
|||
Db 1038 CCAGTAAATATAAATCATGCTCTAGTGGTTCCTTCAGAGAATCTGTGACGCTCAAGAG 1097
QY 123 GlnIleHisProSerArgLeuArgSerProProGluAlaLeuValGlnGlyArgTyr 142
|||
Db 1098 GAAATCAGAAAGTGGAGAAG-----AGAGCTACTTAT 1130
QY 143 ProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLysMetGlu 162
|||
Db 1131 GTTCAT---AAGATGGACTAAATTCACATGATCATGTCGCGGACACTGAAAGTTAT 1187
QY 163 AsnCysLeuGlyGluSerArgHisGluValGluLysSerGluIleSerGluAsnThrAsp 182
|||
Db 1188 GATGTCAGTGAATCATCCGCAAGTTCGAGTGCCTCCTCCCTGTGAGACACACACAG 1247
QY 183 -----AlaSerGlyLysIleGluLysTyrAsn 191
|||
Db 1248 AGATATGAAGGGCCCAACCGAACTGTTCAAAATGGCTGAAATTTCTGTAATGACCCCTGAA 1307
QY 192 ValProLeuAsnArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIle 211
|||
Db 1308 AATGAATAAACAGATGTTTCAGGAATTCAGGATGTCGCCCATGTCGAGCAAGTCA 1367
QY 212 ---LeuArgAlaGlnSerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSer 230
|||
Db 1368 AATGAAGAGTTCATGCAAGGGAGAAACAAACCATACATACACAGAAAGTCTGTACA 1427
QY 231 LeuAspAspLeuGluIleGlyProGlyGlnLeuSerSerSerThrPheAspSerGluLys 250
|||
Db 1428 TTTTGTAAAGGAGGAATTTGGATTAACTCTTTAGGAAACACAGCTTTTACAGACTTTTCT 1487
QY 251 AsnGluSerArgArgAsnLeu-----GluLeuProArgLeu---Ser 263
|||
Db 1488 TGCAACATCTCAGAACTCCGAGAAAGATTCCTGTTAAGACAGCCCGGATCTGCTCT 1547
QY 264 GluThr---SerIleLysAspArgMetAla-----LysTyrGlnAlaAla 277
|||
Db 1548 GAAACAGGTCTCTAAGTGAACATTTCTCAGGCATGATGATGATTTGAGAGTCAAAATGTT 1607
QY 278 ValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGlu 297
|||
Db 1608 GAGTCGAAGATGAAACCTCTTTCATCATAGCTCAGAGTGCAGAAATCTGCGTGTGAC 1667
QY 298 IleLysIleHisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIle 317
|||
Db 1668 TTCAGCTT-----GCCCAACCAACCTATGAGGATGTCATT 1703
QY 318 ThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThr 337
|||
Db 1704 GCT-----GGACATATTTAGATATCTCTGATTCCTTAAGAAAGTAAGAAAAAT 1754
QY 338 -----ProAlaGluAsp 341

Db 1755 TTTCAAAGACGTCGCAAGAGAGTGAAGAGTGTAAAAAGCCCTGGGATATGCAACCGCA 1814
QY 342 AspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysProLeu 361
|||
Db 1815 GATGCTTCTCAACTGAGATGAGAACCCCTTCCAAGAG----- 1853
QY 362 SerProAspSerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLysLys 381
|||
Db 1854 -----GAATCTGCATTTATAAGTGAAGCTGCT----- 1880
QY 382 PheGlnAlaProAla-ArgGluThrCysValGluCysGlnLysThrValTyrProMetGlu 401
|||
Db 1881 -----GCTCCAAGACAAGAAATATGTATCTTTGTCAAAAGACAGTTTATCCAATGGA 1934
QY 401 uArgLeuLeuAlaAsnGlnValPheHisIleSerCysPheArgCysSerTyrCysAs 421
|||
Db 1935 GTCCCTAGTGGCAGACAAGCAGAAATTTTCATAAGTCTGCTCCGATGCCACCATTCGCA 1994
QY 421 nAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgLysCysLysProHis 441
|||
Db 1995 CAGTAACTAAGTTTGGGAAATTCATCATCTTCATGGACAAATATATCTGTAAACCTCA 2054
QY 441 sPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgProHis 461
|||
Db 2055 CTTTAAACAACCTTTTCAAATCCAAAGGAATTTATGATGAAGGTTTTTGGACATAAGCAGCA 2114
QY 461 sLysAspLeuTrpAlaSerLysAsnGluAsnGluGluIle-----LeuGluAr 477
|||
Db 2115 TAAAGATAGATGGAACCTGCAAAACCAAGCAGATCAGTGGACTTTTATCTTCAATCAAGA 2174
QY 477 gProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspLysPr 497
|||
Db 2175 ACCAAATATGTGTAAATATATTGCAGAAACACCCCTGTACCTGGAGATCGTAATAAGCA 2234
QY 497 oIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnG 517
|||
Db 2235 TTTA-----GATGCTGTGTAACAGTGAAGGGCA 2261
QY 517 uLysGluAsp-----LysProAlaGluThrLysLysLeuArgIleAlaTrpProPro 535
|||
Db 2262 AAGGAATGATTTGAGAAATTTAGGGGAAAGGGGAAATTTAAAGTCATTTGGCCCTCTTC 2321
QY 535 oThrGluLeuGlySerSerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLys 555
|||
Db 2322 CAAGGAGATCCCTTAAAGAAACCTTACCTTTTGGAGAGAGCTCAAAATGAGTAAACCTAA 2381
QY 555 sTrpProProGlu---AspGluIleSerLysProGluValProGluAspValAspLeuAs 574
|||
Db 2382 GTGGCCACTGAAATGACAAACCTGCTATCCCTGAAATTTAAAGTGAATCTCTCTCTAGA 2441
QY 574 pLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPheThrValAl 594
|||
Db 2442 AGATGTTAGAACTCCAGAAATATAAGGACAAAGACAAAGATCACTTTCCATTT---TTGCA 2498
QY 594 alaSerPheGlnSerThrSerValLysSerProLysThrValSerProProIleArgly 614
|||
Db 2499 GCCTTATCTACAGTCCACCCATGTTGT----- 2526
QY 614 sGlyTrpSerMetSerGluGlnSerGluGluSerValGlyArgValAlaGluArgly 634
|||
Db 2527 -----CAGAAAGAGGATGTTATAGGA-----ATCAAGAAAGAAATGAA 2561
QY 634 sGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrTrpGlu 654
|||
Db 2562 AATGCTTGAAGAAAGAAAGATGAAAGAAAGAA-----GGAAGGAAAGATGTGCA 2612
QY 654 nAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMe 674
|||
Db 2613 AGATAGCCGAGTGAAGCTGAAGACACAAAGAGTAACAGGAAAGTGTCTGATCTCTTAA 2672
QY 674 tGluAsnGluAsnLeuValGluAsnGlyAlaAsp 685
|||

Db 2673 TGACACAAATAATGTGTAATGTGCAGAGTGTCTGAA 2706

RESULT 10
 US-09-833-381-1727
 ; Sequence 1727, Application US/09833381
 ; Patent No. US20020132090A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Robison, Keith E.
 ; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
 ; FILE REFERENCE: 5800-119
 ; CURRENT APPLICATION NUMBER: US/09/833,381
 ; CURRENT FILING DATE: 2001-04-11
 ; PRIOR APPLICATION NUMBER: 09/516,448
 ; PRIOR FILING DATE: 2000-02-29
 ; NUMBER OF SEQ ID NOS: 2050
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1727
 ; LENGTH: 3236
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(3236)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-833-381-1727

Alignment Scores:
 Pred. No.: 1,118-45 Length: 3236
 Score: 608.00 Matches: 195
 Percent Similarity: 42.88% Conservativity: 112
 Best Local Similarity: 27.23% Mismatches: 260
 Query Match: 15.48% Indels: 149
 DB: 10 Gaps: 25

US-09-890-549-4 (1-759) x US-09-833-381-1727 (1-3236)

QY 51 GluLysLysArgSerAsnThrGluAsnLeuSerGlnHisPheArgLysGlyThrLeuThr 70
 DB 113 GAAATTCACAGAGCAAACTTCCCTTCTCCACCAGGAGTCGCTGGAACAACCTTGT 172
 QY 71 ValLeuLysLysLysTrpGluAsnProGlyLeuGlyAla 83
 DB 173 AGACTCAAGACACCACCTGCAAGTATCCAAAGGGGCCATCCCATGTCCAGCAGCAACC 232
 QY 84 -----GluSerHisThrAspSerLeuArgAsnSerThr----- 95
 DB 233 CCGGTTCCAAATTTAGACAGAGGCTGAAATCATCATGTCTCTCGCAACACACTTCGTCGT 292
 QY 96 -----GluLeuArgHisArgAlaAspHisProProAlaGluValThr----- 109
 DB 293 CAATTAAGATAGAAACTCTGGTGGAGCTCTCCACCTACAATCACAATACCAAGTAAAT 352
 QY 110 ---SerHisAlaAlaSerGlyAla-----LysAlaAspGlnGluGlnIleHis 125
 DB 353 ATAAATCATGCTGCTAGTGGTTCCTTCAGAGAACTCTGGAGCGCTCAAGGAAATCAGG 412
 QY 126 ProArgSerArgLeuArgSerProProGluAlaLeuValGlnGlyArgTyrProHisIle 145
 DB 413 AAAGTGGAGAAG-----AAAGTACTTATGTTTCAT--- 442
 QY 146 LysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLysMetGluAsn:CysLeu 165
 DB 443 AAAGATGCATTAATTCACCTGATCATCATGGTGGCCGACACTGAAAGTTATGATGCAGTT 502
 QY 166 GlyGluSerArgHisGluValGluLysSerGluIleSerGluAsnThrAsp----- 182
 DB 503 GAAATCATCCCAAGGTTCGATGCTCCCTCGCTGCTGACAGACACACACAGATATGAA 562
 QY 183 -----AlaSerGlyLysIleGluLysTyrAsnValProLeu 194
 DB 563 GCGGCCAACCGAAGCTGTTCAATCGCTGCAAAATTCGTGAATGACCTGAAATGAAATA 622
 QY 195 AsnArgLeuLysMetMetPheGluLysGlyGluProThrGlnThrLysIle---LeuArg 213

Db 523 AACAGATGGTTCAGGGAATTTGACATGGCCAGTTTCTGAAGCAAAAGTCAAAATAGAAGA 682
 QY 214 AlaGlnSerArgSerAlaSerGlyArgLysIleSerGluAsnSerTyrSerLeuAspAsp 233
 DB 683 GTTTATGCAAGGGAGAGAAACAAACATAACAAAGAAAGTCTGTACATTTTGTAAAG 742
 QY 234 LeuGluLeuGlyProGlyGlnLeuSerSerThrPheAspSerGluLysAsnGluSer 253
 DB 743 GAGGAATTTGGATTAAACATCTTTAGGAAACACGAGTTTACAGACTTTCTTTCGCAACAT 802
 QY 254 ArgArgAsnLeu-----GluLeuProArgLeu---SerGluThr--- 265
 DB 803 CTTAGAGAACTGGCAGAAAAGATTCCTGTTTAAGCAGCCAGGATCTGCTCTGAACACAG 862
 QY 266 SerIleLysAspArgMetAla-----LysTyrGlnAlaAlaValSerLys 280
 DB 863 TCTTAAGTGAACATTTCTCAAGCATGGATGCAATTTGAGAGTCAATTTGTTGATGCTGAAG 922
 QY 281 GlnSerSerThrAsnTyrThrAsnGluLeuLysAlaSerGlyGlyGluIleLysIle 300
 DB 923 ATGAAAACCTCTTTCATCATAGTCTGAGAGCTGGCAAAATCTGGCTGTGACTTCAAG--- 979
 QY 301 HisLysMetGluGlnLysGluAsnValProProGlyProGluValCysIleThrHisGln 320
 DB 980 -----CATGCCCAACCACTATGAGGATGTCTTGTCT----- 1012
 QY 321 GluGlyGluLysIleSerAlaAsnGluAsnSerLeuAlaValArgSerThr----- 337
 DB 1013 ---GGACATATTTTAGATATCTCTGATTCACCTTAAAGAAAGTAAAGAAAAATTTTCAAAG 1069
 QY 338 -----ProAlaGluAspAspSerArg 344
 DB 1070 ACGTGGCAAGAGAGTGAAGAGTTTTTAAGGGCTGGGATATGCAACCCGAGATGTCTTCT 1129
 QY 345 AspSerGlnValLysSerGluValGlnGlnProValHisProLysProLeuSerProAsp 364
 DB 1130 GCAACTGAGATGAAGAACCTTCCCAAGAG----- 1159
 QY 365 SerArgAlaSerSerLeuSerGluSerSerProProLysAlaMetLysLysPheGlnAla 384
 DB 1160 ---GAATCTGCATTATATAAGTGAAGCTGCT-----GCT 1189
 QY 385 ProAla-ArgGluThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLe 404
 DB 1190 CCAAGACAGGAAATATGATATCTTGTCAAAGACAGTTATTCCTCAATGGAGTCCCTAGT 1249
 QY 404 uAlaAsnGlnGlnValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLysLe 424
 DB 1250 GGCAGACAAGCAGAAATTTTCATAAGTCTCTGCTCCGATGCCACCATGCAACAGTAACT 1309
 QY 424 uSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnG 444
 DB 1310 AAGTTTGGGAAATATGTCATCTCATGGACAAATATACTGTAAACCTCCTTTAAACA 1369
 QY 444 nLeuPheLysSerLysGlyAsnTyrAspGluClyPheGlyHisArgProHisLysAspLe 464
 DB 1370 ACTTTTCAAAATCCAAAGGAAATTTATGATGAAGTTTTCGACATTAAGCAGCATAAAGATAG 1429
 QY 464 uTrpAlaSerLysAsnGluAsnGluIle-----LeuGluArgProAlaG 480
 DB 1430 ATGGAACTCCAAAACCAAGCAGATCAGTGGACTTTTATTCCTTAATGAAGAACCAATAT 1489
 QY 480 nLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaProIleAlaL 500
 DB 1490 GTGTAATAATATTGAGAAAAACACCTTGTCTCTGGAGATCGTAAATGAACATTTA----- 1544
 QY 500 sValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAs 520
 DB 1545 -----GATGCTCGTAAACAGTGAAGGCAAGGCAAGCAATGA 1576
 QY 520 p-----LysProAlaGluThrLysLysLysLeuArgIleAlaTrpProProThrGluLe 538

Db 1577 TTGAGAAAATTAGGGGAAAGGGGAAATTTAAAGTCAATTTGGCCCTCTTCCCAAGGAGAT 1636
Qy 538 uGlySerSerGlySerAlaLeuGluGluGlyIleLysMetSerLysProLysTrpProR 558
Db 1637 CCCTAAGAAAACCTTACCCTTTGAGGAGAGCTCAAAATGATGATAAAGTGGCCACC 1696
Qy 558 oGlu---AspGluIleSerLysProGluValProGluAspValAspLeuAspLeuLys 577
Db 1697 TGAATGACAAACCTGCTATCCCTGATTTAAAGTGAATCTCTGCTAGAGATTTAG 1756
Qy 577 sLeuArgSerSerSerLeuLysGluArgSerA:GProPheThrValAlaAlaSerPh 597
Db 1757 AACTCCAGAAATAAAGGACAAAGACAAAGATCACTTTCCATTT---TTGCAGCCTTATCT 1813
Qy 597 eGlnSerThrSerValLysSerProLysThrValSerProLysArgLysGlyTrpSe 617
Db 1814 ACAGTCACCCATGTTTGT----- 1832
Qy 617 rMetSerGluGlnSerGluGluSerValGlyGlyArgValAlaGluArgLysGlnValG 637
Db 1833 -----CAGAAGAGAGATCTTATAGGA-----ATCAAGAAATGAAATGCTCGA 1876
Qy 637 uAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrTrpGlnAsnLysG 657
Db 1877 AGAAGAAAGATGAAAGAAAGAA-----GGAAGGAAGATGTCAAGATAGGCC 1927
Qy 657 uSerLysGlyGluThrGlyLysArgSerLysGluGlyHisSerLeuGluMetGluAsnG 677
Db 1928 GAGTGAAGCTGAAGACACAAAGAGTAAACAGGAAAGTGTCTATGATCTTTAATGACAA 1987
Qy 677 uAsnLeuGluAsnGlnValAlaAspSerAspGluAspAspAsnSer 692
Db 1988 TAATGTGATGTGCGAGTGTCTGAAAGAGGAGAAATGAAAAAACT 2033

RESULT 11

US-09-909-567B-2
; Sequence 2, Application US/0909567B
; Publication No. US20030022257A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto A.
; APPLICANT: Nair, Manoj
; APPLICANT: Chen, Seiyu
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes
; FILE REFERENCE: DEX-0214
; CURRENT APPLICATION NUMBER: US/09/909,567B
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 60/219,834
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 3825
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (429)..(428)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (430)..(430)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1997)..(1998)
; OTHER INFORMATION: a, c, g, or t
US-09-909-567B-2

Alignment Scores:
Pred. No.: 1.58e-44 Length: 3825
Score: 596.50 Matches: 203
Percent Similarity: 42.50% Conservative: 120
Best Local Similarity: 26.71% Mismatches: 282

Query Match: 15.19% Indels: 155
DB: 11 Gaps: 27
US-09-890-549-4 (1-759) x US-09-909-567B-2 (1-3825)
Qy 12 ThrSerLeuSerLeuArgValThrAlaLysGluLeuSerLeuValAsnLysAsnLysSer 31
Db 87 ACTTTTATCAAAATAGAAATCTACTGCCCGCAGAAACAGAAAAACCCCTACTTAAGAACGAGCTT 146
Qy 32 SerAlaIleValGluIlePheSerLysLysTrpGlnLysAlaAlaGluGluThrAsnMetGlu 51
Db 147 TCTCAGTCCCTCTAAAGGACAGT---TATGTTGAAACCCCAAGAGGCCCATGTCTG 203
Qy 52 LysLys-----ArgSerAsnThrGluAsnLeuSerGlnHisPheArgLysGly 67
Db 204 CAAAAATCTGAAATTCACAGAGCAACACTTCCCTCTCTCCACCCAGGAGTCGCTCTGAA 263
Qy 68 ThrLeuThrValLeuLysLysLysTrpGluAsnProGlyLeuGlyAla----- 83
Db 264 CAACCTGTGACAGACTCAAAAGACACCACTGCAAAAGTTATCCAAAGGGGCCATCCCATGTCCA 323
Qy 84 -----GluSerHisThrAspSerLeuArgAsnSerSerThr 95
Db 324 CGAGCAACCCCGTTCCTCAATTTGTAGAGAGAGGCTTGAATATCATCATGTCTCTCTGCAACA 383
Qy 96 -----GluIleArgHisArgAlaAspHisProPro-AlaGluValThr-- 109
Db 384 CTTCGTGTCCTCAATTAAGATAGAACTCGTGGTAGGACTCTCCNANCTACATCAATCAAT 443
Qy 110 -----SerHisAlaLysSerGlyAla-----LysAlaAspGlnGlu 122
Db 444 ACCAGTAATATAAATCATGCTGTAGTGTCTCTCCAGAGAAATCTGTGGACGCTCAAGA 503
Qy 122 uGlnIleHisProArgSerArgLeuArgSerProProGluAlaLeuValGlnGlyArgTy 142
Db 504 GGAATCAGGAAGCTGGAGAG-----AGAGCTACTTA 536
Qy 142 rProHisIleLysAspGlyGluAspLeuLysAspHisSerThrGluSerLysLysMetG 162
Db 537 TGTTCTAT---AAGATGGACTAAATCCACTGATCATGTGTCGCCGACACTGAAAGTTA 593
Qy 162 uAsnCysLeuGlyLysSerArgHisGluValGluLysSerGluIleSerGluAsnThrAs 182
Db 594 TGATGCAGTTGAAATCATCCGCAAGGTTGCAAGTTCCTCGCTGTCTGAGGACACACACA 633
Qy 182 p-----AlaSerGlyLysIleGluLysTyAs 191
Db 654 GAGATATGAAGCGCCCAACCGAACTGTTCAAATGGCTGMAAAATTCGTGAATCACCCTGA 713
Qy 191 nValProLeuAsnArgLeuLysMetPheGluLysGlyGluProThrGlnThrLysI 211
Db 714 AAATGAAATAAACAGATGTTTACGGGAAATTTGAGCATGGCCCAAGTTTCTGAAGCAAAAGTC 773
Qy 211 e---LeuArgAlaGlnSerArgSerAlaSerGlyArgLysIleSerGluAsnSerTySe 230
Db 774 AAATAGAGAGAGTTTATGCAAGGAGGAGAAACCAACATAACATAACAACAAGAGTCGTAT 833
Qy 230 rLeuAspAspLeuGluIleGlyProGlyGlnLeuSerSerSerThrPheAspSerGlu 250
Db 834 ATTTTGAAGGAGGAATTTGGATTAACATCTTTAGGAAACACGAGTTTACAGACTTTTC 893
Qy 250 sAsnGluSerArgArgAsnLeu-----GluLeuProArgLeu---Se 263
Db 894 TTGCAAAACATCTTAGAGAACTCGGAAAAAGATTTCCTGTTAAGCAGCCCGAGGATCTGCTC 953
Qy 263 rGluThr---SerIleLysAspArgMetAla-----LysTyGlnAla 277
Db 954 TGAACACAGGTCTCTAAGTGAACATTTCTCAGGCGATGGATGATTTGAGAGTCAATTTGT 1013
Qy 277 aValSerLysGlnSerSerSerThrAsnTyThrAsnGluLeuLysAlaSerGlyGly 297
Db 1014 TGAGTCGAAGATGAAACCTCTTTCATCATCATAGTCTCAGAGCTGCAAAATCTGCTGTGA 1073

QY 297 uilelyslleHisLysMetGluGlnLysGluAsnValProProGlyProGluValCysI 317
Db 1074 CTTCAAG-----CATGCCCAACCACTATGAGGTGTCAT 1109
QY 317 efrHisGlnGluGlyGluLysLysSerAlaAsnGluAsnSerLeuAlaValArgSerTh 337
Db 1110 TGCT-----GGACATATTTAGATATCTCTGATTACCTTAAGAGTAAGAAAAA 1160
QY 337 r-----ProAlaGluAs 341
Db 1161 TTTTCAABAGCGTGGCAAGAGAGTGGAGAGTGTTTTAAAGCGCTGGGATGCAACCGC 1220
QY 341 pAspSerArgAspSerGlnValLysSerGluValGlnGlnProValHisProLysProLe 361
Db 1221 AGATGCTTCTGCACTGAGATGAGAACCACTTCCAGAG----- 1260
QY 361 userProAspSerArgAlaSerSerLeuSerGluLysSerSerProProLysAlaMetLysLy 381
Db 1261 -----GAATCTGCATTATTAAGTGAAGCTGCT----- 1287
QY 381 sPheGlnAlaProAla-ArgGluThrCysValGluCysGlnLysThrValTyProMetG 401
Db 1288 -----GCTCCAGACAGGAATATGTATCTTTGTCAAAAGACAGATTATCCAAATGG 1340
QY 401 luArgLeuLeuAlaAsnGlnGlnValPheHisLysSerCysPheArgCysSerTyrCysA 421
Db 1341 AGTGCTAGTGGCAGACAGCAGAAATTTTCATAGTCTCTGCTCCGATGCCACCATTTGCA 1400
QY 421 snAsnLysLeuSerLeuGlyThrTyrAlaSerLeuHisGlyArgIleTyrCysLysProH 441
Db 1401 ACAGTAACTAAGTTTGGGAAATTTATGCATCTTCATGACAAATATCTGTAACCTC 1460
QY 441 isPheAsnGlnLeuPheLysSerLysGlyAsnTyrAspGluGlyPheGlyHisArgProH 461
Db 1461 ACTTTAAACAACCTTTTCAAAATCCAAAGGAAATATGATGAAGTTTGGACATACAGC 1520
QY 461 isLysAspLeuTrpAlaSerLysAsnGluAsnGluIle-----LeuGluA 477
Db 1521 ATAAAGATAGTGAACCTGCAAAACCAAGACAGATCAGTGGACTTTATTCCTTAATGAAG 1580
QY 477 rGProAlaGlnLeuAlaAsnAlaArgGluThrProHisSerProGlyValGluAspAlaP 497
Db 1581 AACCAAAATATGTGTAAATAATTTGCAGAAACACCCCTTGACCTGGAGATCGTAATGAAC 1640
QY 497 roIleAlaLysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnG 517
Db 1641 ATTTA-----GATGCTGGTAAACAGTGAAGGCG 1667
QY 517 lLysGluAsp-----LysProAlaGluThrLysLysLeuArgIleAlaTrpProProp 535
Db 1668 AAGGAATGATTTGAGAAATTTAGGGGAAAGGGGAAATTTAAAGTCATTTGGCCTCCTT 1727
QY 535 roThrGluLeuGlySerSerGlySerAlaLeuGluGlyLysMetSerLysProL 555
Db 1728 CCAAGGAGATCCCTAAGAAACCTTACCCTTTGAGGAGAGCTCAAAATGATGAACCTTA 1787
QY 555 ystrProProGlu-----AspGluLysSerLysProGluValProGluAspValAspLeuA 574
Db 1788 AGTGGCCACCTGAAATGACAAACCTCTATCCCTGAATTTAAAGTGAATCTCTGCTAG 1847
QY 574 sPLeuLysLysLeuArgArgSerSerLeuLysGluArgSerArgProPheThrValA 594
Db 1848 AGATGTTAGAACTCCAGAAATAAAGGCAAGACAGATCACTTTCCATTT---TTGC 1904
QY 594 laAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProProfileArgL 614
Db 1905 AGCCTTATCTACAGTCCACCATGTTGT----- 1933
QY 614 ysglyTrpSerMetSerGluGlnSerGluSerValGlyGlyArgValAlaGluAgl 634
Db 1934 -----CAGAAAGGATGTTATAGGA-----ATCAAGAAATGA 1967
QY 634 ysglnValGluAsnAlaLysAlaSerLysLysAsnGlyAsnValGlyLysThrTrpG 654

Db 1968 AATGCTGAGGAGAGAAAGATGAAGNNGAA-----GGAAGGAGAAATGTGC 2018
QY 654 lnAsnLysGluSerLysGlyGluThrGlyLysArgSerLysGlyHisSerLeuGluM 674
Db 2019 AAGATAGGCGGAGTGAAGCTGAAGACACACAGAGTAAACAGGAAAGTCTATGATCTTA 2078
QY 674 etGluAsnGluAsnLeuValGluAsnGlyAlaAspSerAspGluAspAsnSer 692
Db 2079 ATGACACAATAATGTGATTGTGCAGAGTGTCTGAAAGGAGAGAAATAATGAAAAAACT 2134

RESULT 12

US-10-161-927-73
; Sequence 73, Application US/10161927
; Publication No. US20030235821A1
; GENERAL INFORMATION:

; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Miller, Charles E.
; APPLICANT: Hjal, Tord
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Vernet, Corine
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Li, Li
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Gorman, Linda
; APPLICANT: Anderson, David W.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Patturajan, Meera
; APPLICANT: Stone, David J.

; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 21402-377 D (Cura 677 Other)
; CURRENT APPLICATION NUMBER: US/10/161,927
; CURRENT FILING DATE: 2002-06-03

; PRIOR APPLICATION NUMBER: 60/295,661
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/295,607
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/296,404
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/296,418
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/296,575
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/297,414
; PRIOR FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: 60/297,567
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/298,528
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/325,685
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/299,133
; PRIOR FILING DATE: 2001-06-18
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 190
; SEQ ID NO 73

; LENGTH: 2955

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1180) .. (2398)

US-10-161-927-73

Alignment Scores:

Pred. No.: 4,37e-40 Length: 2955


```

; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Valda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/09/880,192
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 13
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020077470A1 058201CB1
US-09-880-192-13

Alignment Scores:
Pred. No.: 4,11e-39 Length: 2379
Score: 534.50 Matches: 149
Percent Similarity: 50.22% Conservative: 77
Best Local Similarity: 33.11% Mismatches: 157
Query Match: 13.61% Indels: 69
DB: 9 Gaps: 16

US-09-890-549-4 (1-759) x US-09-880-192-13 (1-2379)

QY 260 ProArgLeu---SerGluThr---SerIleYsAspArgMetAla----- 272
Db 1 CCCAGGATCTCTCTGAACACAGCTCTCTAAGTGAACATCTTCAGGCATGGATGCAATT 60

QY 273 LysTyrGlnAlaAlaValSerLysGlnSerSerSerThrAsnTyrThrAsnGluLeuLys 292
Db 61 GAGAGTCATATTGTGAGTCGACATGAACATGAACCTCTTCATCATCATAGCTCAGAGCTGC 120

QY 293 AlaSerGlyGluIleLysIleHisLysMetGluGlnLysGluAsnValProProGly 312
Db 121 AAATCTGGCTGTGACTTCAAG-----CATGCCCCACCAACC 156

QY 313 ProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeu 332
Db 157 TATGAGGATGTCATTGCT-----GGACATATTTAGATATCTCTGATTCACCTAAA 207

QY 333 AlaValArgSerThrProAlaGluAspAspSerArgAsp-----SerGlnValLysSer 350
Db 208 GAAGTAAGAAAAAATTT-CAAAAGACGCTGGCAAGAGAGTGAAGAGATT-TTAAAGGCC 265

QY 351 GluValGlnGlnProValHisProLysProLysProLysSerProAspSerArgAlaSerSerLeu 370
Db 266 TGGATATGCAACCG-----CAGATGCTTCTGCAACATGAGATGAGAACCCTTCCAA 319

QY 371 SerGluSerSerProLysAlaMetLysLysPheGlnAlaProAla-ArgGluThrCy 390
Db 320 GAGGAATCTGCA-----TTTAAAGTGAAGCTGCTCTCCAGACAGGAATATG 370

QY 390 sValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGlnValPh 410
Db 371 TATCTTTGTCAAAAGACAGATTATCAATGGAGTGCCTAGTGGCAGACAAGCAGAAATT 430

QY 410 eHisIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGluThrTyAl 430
Db 431 TCATAAGTCTGCTCCGATGTCACCATTTGCAACAGATAAAGTAAAGTTTGGAAATATG 490

QY 430 aserLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLysSerLysGl 450
Db 491 ATCATTTCATGACAAATATATCTGTAACCTCACITTAACACACTTTTCAATCCAAAG 550

QY 450 YAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSerLysAsnGl 470
Db 551 AAATTATGATGAGGTTTGGACATGAAGCAGATGAAGATAGATGCAAGCTGCAAAACCA 610

QY 470 uAsnGluGluIle-----LeuGluArgProAlaGlnLeuAlaAsnAlaArgGl 486

```

```

Db 611 AAGCAGATCAGTGCAGCTTTATTCTTAATGAAGAACCAATATATGTGTAAATAATTTCAGA 670
QY 486 uThrProHisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAlaAl 506
Db 671 AAACACCTTTGTACTGGAGATCGTAATGAACATTTA----- 707
QY 506 aSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp-----LysProAlaGl 524
Db 708 -----GATGCTGGTAACAGTGAAGGCAAGCAAGCAATGATTTGAGAAAATTAGGGGA 757
QY 524 uThrLysLysLeuArgIleAlaTyrProProThrGluLeuGluGlySerSerGlySerAl 544
Db 758 AAGGGGAAAATTAAAGTCAATTTGGCTCTCTTCCAAGAGATCCCTTAAGAAAAACCTTACC 817
QY 544 aLeuGluGluGlyIleLysMetSerLysProLysTyrProProGlu-----AspGluIleSe 563
Db 818 CTTTGAGGAAGAGCTCAAAATGAGTAAACCTAAGTGGCCACCTGAAATGACAAACCTGCT 877
QY 563 rLysProGluValProGluAspValAspLeuAspLeuLysLysLeuArgArgSerSerSe 583
Db 878 ATCCCTGAATTTAAAGTGAATCTCTGCTAGAAGATGTTAGAACTCCAGAAAAATAAAGG 937
QY 583 rLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThrSerVally 603
Db 938 ACAAGACCAAGATCACTTTCCATTT---TTGAGGCTTATCTACAGTCCACCCATGTTTG 994
QY 603 sSerProLysThrValSerProProIleArgLysGlyTyrSerMetSerGluGlnSerGl 623
Db 995 T-----CAGAAAGA 1003
QY 623 uGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSerly 643
Db 1004 CGATGTTATAGA-----ATCAAGAAATGAAATGCTCTGAAGGAAGAAAGATGAANA 1057
QY 643 sLysAsnGlyAsnValGlyLysThrThrTrpGlnAsnLysGluSerLysGlyGluThrGl 663
Db 1058 GAAGAA-----GGAAGGAAGATGCTCAAGATAGGCGGAGCTGAAGCTGAAGACAC 1108
QY 663 yLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnGl 683
Db 1109 AAAGAGTAAAGGAAAGTGTATGATCTTAATGACAACAATAATGTGATTTGTCAGAG 1168
QY 683 yAlaAspSerSerGluAspAspAsnSer 692
Db 1169 TGCTGAAAGGAGAAATGAAAAAAT 1196

RESULT 14
US-10-427-348-13
; Sequence 13, Application US/10427348
; Publication No. US20030175795A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Valda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-2 CON
; CURRENT APPLICATION NUMBER: US/10/427,348
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/880,192
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: US 09/299,708
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 13
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175795A1 058201CB1

```

```

US-10-427-348-13
Alignment Scores:
Pred. No.: 4,11e-39 Length: 2379
Score: 534.50 Matches: 149
Percent Similarity: 50.22% Conservative: 77
Best Local Similarity: 33.11% Mismatches: 157
Query Match: 13.61% Indels: 69
DB: 13 Gaps: 16

US-09-890-549-4 (1-759) x US-10-427-348-13 (1-2379)
QY 260 ProArgLeu---SerGluThr---SerIleLysAspArgMetAla----- 272
Db 1 CCCAGGATGCTCTCTGAACACAGGCTCTTAAGTGAACATTTCTCAGGCATGGATGCAATT 60
QY 273 LysTyGlnAlaAlaValSerLysGlnSerSerThrAsnTyThrAsnGluLeuLys 292
Db 61 GAGAGTCAAAATTTGTCGAGTGAAGATGAACCTCTTCATCATAGCTCAGAGCTGCC 120
QY 293 AlaSerGlyGlyGluIleLysLysMetGluGlnLysGluAsnValProProGly 312
Db 121 AAATCTGGCTGTGACTTCAAG-----CATGCCCCACCAACC 156
QY 313 ProGluValCysIleThrHisGlnGluGlyGluLysIleSerAlaAsnGluAsnSerLeu 332
Db 157 TATGAGATGCTATTGCT-----GGACATATTTAGATATCTCTGATTCACCTAAA 207
QY 333 AlaValArgSerThrProAlaGluAspAspArgAsp-----SerGlnValLysSer 350
Db 208 GAAGTAAGAAAAAATTT-TCAAAAGACGTGGCAAGAGAGTGGAGAGTGT-TTAAGGCC 265
QY 351 GluValGlnGlnProValHisProLysProLysProLysProLysProLysSerLeu 370
Db 266 TGGGATATGCAACCG-----CAGATGCTTCTGCAACATGAGATGAGAACCACTTCCAA 319
QY 371 SerGluSerSerProLysAlaMetLysLysPheGlnAlaProAla-ArgGluThrCy 390
Db 320 GAGGAATCTGCA-----TTTATAGTGAAGCTGCTGCCAAGCAAGAAATATG 370
QY 390 sValGluCysGlnLysThrValTyProMetGluArgLeuLeuAlaAsnGlnValph 410
Db 371 TATACTTTGTCAAAAGACAGTTTATCCATGGAGTGCTAGTGGCAGACAGCAATTT 430
QY 410 eHisIleSerCysPheArgCysSerTyCysAsnAsnLysLeuSerLeuGlyThrAl 430
Db 431 TCATAAGTCTCTGCTCCGATGCCACCATTCACCAAGTAACTAAGTTTGGCAATTTATGC 490
QY 430 aSerLeuHisGlyArgIleTyCysLysProHisPheAsnGlnLeuPheLysSerLysG 450
Db 491 ATCACTTTATGGACAAATATATCTGTAACCTTCAACCTTTTAAACAACCTTTTCAAAATCCAAAGG 550
QY 450 yAsnTyArgGluGlyPheGlyHisArgProHisLysAspLeuTrpAlaSerLysAsnG 470
Db 551 AAATATATGATGAAGGTTTGGACATAAGCAGCATAAAGATAGATGGAAGTCCAAAACCA 610
QY 470 uAsnGluGlulle-----LeuGluArgProAlaGlnLeuAlaAsnAlaArgL 486
Db 611 AAGCAGATCAGTGGACTTTATTCCTAATGAAGAACCAAAATATGTGTAATAATATTCAGA 670
QY 486 uThrProHisSerProGlyValGluAspAlaProIleAlaLysValGlyValLeuAla 506
Db 671 AAACACCTTTGATCTGGAGATCGTAAATGAACATTTA----- 707
QY 506 aSerMetGluAlaLysAlaSerSerGlnGlnGluLysGluAsp-----LysProAlaG 524
Db 708 -----CATGCTGTAACAGTGAAGGCCAAAGGAATGATTTGAGAAAAATAGGGGA 757
QY 524 uThrLysLysLeuArgIleAlaTrpProProProProThrGluLeuGlySerSerAl 544
Db 758 AAGGGGAAAAATTAAGAGTCATTTGGCTCTCTTCCAGGAGATCCCTAAGAAAAACCTTACC 817
QY 544 aLeuGluGluGlyIleLysMetSerLysProLysProLysTrpProGlu---AspGluIleSe 563

```

818 CTTTGAAGAGAGCTCAAAATGAGTAAACCTTAAGTGGCCACCTGAAATGACAAACCTGCT 877
 QY 563 rLysProGluValProGluValAspValAspLeuLysLysLeuArgArgSerSerSe 583
 Db 878 ATCCCTGAAATTTAAAGTGAATCTCTGAGAGATGTTAGAACTCCAGAAAATAAAGG 937
 QY 583 rLeuLysGluArgSerArgProPheThrValAlaAlaSerPheGlnSerThrSerVal 603
 Db 938 ACAAAGACACAGATCATCTTCCATTT---TTGCACCTTATCTACAGTCCACCCATGTTG 994
 QY 603 sSerProLysThrValSerProIleLysGlyTrpSerMetSerGluGlnSerG 623
 Db 995 T-----CAGAAAGA 1003
 QY 623 uGluSerValGlyGlyArgValAlaGluArgLysGlnValGluAsnAlaLysAlaSer 643
 Db 1004 GGATGTTATAGGA-----ATCAAAGAAATGAAATATCCCTGAAAGGAAAGAAATGAAA 1057
 QY 643 sLysAsnGlyAsnValGlyLysThrTrpGlnAsnLysGluSerLysGlyGluThrG 663
 Db 1058 GAAGGAA-----GGAAGGAAGATGTGCAAGATAGCCGAGTGAAGCTGAAGACAC 1108
 QY 663 yLysArgSerLysGluGlyHisSerLeuGluMetGluAsnGluAsnLeuValGluAsnG 683
 Db 1109 AAAGAGTAACAGGAAAGTGTATGGATCTTAATGACAAATAATGTTGTCGAGAG 1168
 QY 683 yAlaAspSerAspGluAspAspAsnSer 692
 Db 1169 TGCTGAAAGAGGAAATAATGAAAAAAT 1196

RESULT 15
 US-10-440-366-1
 ; Sequence 1, Application US/10440366
 ; Publication No. US20030181706A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walker, Michael G.
 ; APPLICANT: Volkmut, Wayne
 ; APPLICANT: Klingler, Tod M.
 ; TITLE OF INVENTION: INFLAMMATION-ASSOCIATED POLYNUCLEOTIDES
 ; FILE REFERENCE: PB-0006-1 CIP
 ; CURRENT APPLICATION NUMBER: US/10/440,366
 ; PRIOR FILING DATE: 2003-05-16
 ; PRIOR APPLICATION NUMBER: US 09/855,323
 ; PRIOR FILING DATE: 2001-05-14
 ; PRIOR APPLICATION NUMBER: US 09/195,292
 ; PRIOR FILING DATE: 1998-11-18
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 1
 ; LENGTH: 1298
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: 402334CB1
 US-10-440-366-1

Alignment Scores:
 Pred. No.: 9,75e-26 Length: 1298
 Score: 385.00 Matches: 122
 Percent Similarity: 42.97% Conservative: 46
 Best Local Similarity: 31.20% Mismatches: 137
 Query Match: 9.80% Indels: 90
 DB: 13 Gaps: 12

US-09-890-549-4 (1-759) x US-10-440-366-1 (1-1298)
 QY 329 GluAsnSerLeuAlaValArgSerThrProAlaGluAspAspSerArgAspSerGlnVal 348
 Db 117 CAGAACCCAGCGGGTCCGCTTCTCCACCC-----GAGGCTTCCACCTCCACGACCA 170
 QY 349 LysSerGluValGlnGlnProValHisProLysProLeu---SerProAspSerArgAla 367

Search completed: January 7, 2004, 01:57:43
Job time : 641 secs

```
Db 171 TGTTCAGGCTGAGGAGCGCCAGGCGACCCCTCTCATGACGCCAAAGCGCGGCA 230
Qy 368 SerSerLeuSerGluSerProProLysAlaMetLysLysPheGlnAlaProAlaArg 387
Db 231 GCAGCAGGTGCAGCGCTCCAGTCCCTCAGCCTG-CGG-----GCCAGGTGAAG 280
Qy 388 GluThrCysValGluCysGlnLysThrValTyrProMetGluArgLeuLeuAlaAsnGln 407
Db 281 GAGACCTGCGCGCGCTGCAGAGACCGTGATCCCATGGAGCGCTGGTGGCCGACAG 340
Qy 408 GlnValPheHisIleSerCysPheArgCysSerTyrCysAsnAsnLysLeuSerLeuGly 427
Db 341 CTCATTTTCCCAACTCTGTCTGTGCAAGCACTGTACACCAAGCTCAGCCTGGGC 400
Qy 428 ThrTyrAlaSerLeuHisGlyArgIleTyrCysLysProHisPheAsnGlnLeuPheLys 447
Db 401 AGCTACGCGCGCTGCAGCGGAGTTCTACTGCAAAACCCCTTCCAGCAGCTGTTAAG 460
Qy 448 SerLysGlyAsnTyrAspGluGlyPheGlyHisArgProHisLysAspLeuTyrAlaSer 467
Db 461 AGCAAGGCAACTACGACGAGGGTTGGCGGCAAGCAGCAGCAGAGGCTCTGGGCCAC 520
Qy 468 LysAsnGluAsnGluGluIleLeuGluArgProAlaGlnLeuAlaAsnAlaArgGluThr 487
Db 521 AAGGAGGTGGAC-----532
Qy 488 ProHisSerProGlyValGluAspAla-----ProlleAla 499
Db 533 -----CCCGCACCAAGACCGCTGAGGCTCTGTAACTTCCACCCCTCTG-- 580
Qy 500 LysValGlyValLeuAlaAlaSerMetGluAlaLysAlaSerSerGlnGlnGluLysGlu 519
Db 581 -----CGGAGGCTTGGAGCGCGGAGGGGGAAGGT 610
Qy 520 AspLysProAlaGluThrLysLysLeuArgIleAlaTyrProProThrGlu----- 537
Db 611 GGAAGGAGGTGCGAG-----CTGGG-CTTGGTGGGGCCAGG-TGGGAAGGGAT 659
Qy 538 -----LeuGlySerSerGlySerAlaLeu-----545
Db 660 GAGCCTTGCTCAGCGGTAGGGGACCGAGGCGGCTCTGCTCAGGACTCCTCTCTT 719
Qy 546 -----GluGluGlyIleLysMetSerLysProLysTyrProProGlu----- 559
Db 720 CTTCTCCCGCAGCGGTGAGGTTGGAAACAGGATTGGGTCTGCCACCCCTGC 779
Qy 560 -----AspGluIleSerLysProGluValProGluAsp-----ValAsp 572
Db 780 TTCTGTCTTCTCAGCCTCCCTCCACCTCACCAGGACCCCTGGAGGCCCCCAA 839
Qy 573 LeuAspLeuLysLysLeuArgArgSerSerSerLeuLysGluArgSerArgProPheThr 592
Db 840 GCCAGCTCCCTCTATAGTGTCTTTCTCCAGCAAGGAGTCAGCATGCCCTCAGG 899
Qy 593 ValAlaAlaSerPheGlnSerThrSerValLysSerProLysThrValSerProPhe 612
Db 900 GTCCAGCTCCCTCTACTGCCACCGGAGA-CTGTGTGGCCCCCACGCTCTCCCATCTACC 958
Qy 613 ArgLysGlyTyrSerMetSerGluGlnSerGluGluSerValGlyGlyArgValAlaGlu 632
Db 959 TCTACCTTAACCTGTTCTGAGCCACGAGAGGAGGAGGAGCGCGCAGTGGCCAC 1018
Qy 633 ArgLysGlnValGluAsnAlaLysAlaSerLysLysAsnGlyAsn-----647
Db 1019 CTGTTGGGCATATAATGCCCTGAGCCCATGGGGAGGAGATGGGGAAGTGGAGCCA 1078
Qy 648 -----ValGlyLysThrThrTyrGlnAsnLysGluSerLysGly 660
Db 1079 CCTGCTCTGAGGCAAGGAGGCGCTGCCCTGAGTGGGCTTGGGACCATCTCGAAC 1138
Qy 661 GluThrGlyLysArgSerLysGluGlyHisSer 671
Db 1139 ACCAGGTGGAGAGCAGAGCAAAAGCACTCG 1171
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:51:46 ; Search time 21 Seconds
(without alignments)
1529.235 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSPNNRROWTSLSLRVTA.....LSVEQIKRNYDEDEEE 759

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.1	398	1	US-08-261-822A-16
2	8	1.1	398	5	PCT-US95-07744A-16
3	8	1.1	582	4	US-08-334-179A-4
4	8	1.1	757	4	US-09-252-991A-20231
5	8	1.1	1038	4	US-08-334-179A-2
6	8	1.1	1038	4	US-08-334-179A-8
7	7	0.9	12	3	US-08-844-978-30
8	7	0.9	12	3	US-08-844-978-33
9	7	0.9	15	4	US-08-979-608A-19
10	7	0.9	25	2	US-08-789-078-15
11	7	0.9	25	2	US-08-752-633-16
12	7	0.9	25	3	US-08-844-978-29
13	7	0.9	25	5	PCT-US95-04886-16
14	7	0.9	26	4	US-08-979-608A-20
15	7	0.9	70	4	US-09-252-991A-27697
16	7	0.9	130	4	US-09-252-991A-26359
17	7	0.9	152	2	US-08-608-143-4
18	7	0.9	156	4	US-09-252-991A-28098
19	7	0.9	159	4	US-09-107-532A-3789
20	7	0.9	180	4	US-09-252-991A-21814
21	7	0.9	189	2	US-08-064-694-2
22	7	0.9	189	4	US-09-198-452A-1169
23	7	0.9	200	3	US-08-881-094-19
24	7	0.9	213	4	US-09-252-991A-20163
25	7	0.9	215	4	US-09-252-991A-21524
26	7	0.9	217	4	US-08-979-608A-7
27	7	0.9	240	4	US-09-107-532A-5420

Sequence 1, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 3, Appli
Sequence 17423, A
Sequence 19816, A
Sequence 372, Appl
Sequence 2, Appli
Sequence 39, Appl
Sequence 371, Appl
Sequence 2, Appli
Sequence 39, Appl
Sequence 21516, A
Sequence 4, Appli
Sequence 322, Appl
Sequence 32185, A
Sequence 4349, Ap

ALIGNMENTS

RESULT 1
US-08-261-822A-16
; Sequence 16, Application US/08261822A
; Patent No. 5650553
; GENERAL INFORMATION:
; APPLICANT: Ecker, Joseph R. et al.
; TITLE OF INVENTION: Plant Genes for Sensitivity to Ethylene
; TITLE OF INVENTION: and Pathogens
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5650553ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,822A
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-261-822A-16

Query Match 1.1%; Score 8; DB 1; Length 398;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 NNKLSLGT 428

Db 215 NNKLSLGT 222

```
RESULT 2
PCT-US95-07744A-16
; Sequence 16, Application PC/TUS9507744A
; GENERAL INFORMATION:
; APPLICANT: Trustees of The University of Pennsylvania
; TITLE OF INVENTION: Plant Genes for Sensitivity to Ethylene
; TITLE OF INVENTION: and Pathogens
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07744A
; FILING DATE: 15-JUNE-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/261,822
; FILING DATE: June 17, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-07744A-16

Query Match 1.1%; Score 8; DB 5; Length 398;
Best Local Similarity 100.0%; Pred.No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 421 NNKLSLGT 428
Db 215 NNKLSLGT 222

RESULT 3
US-08-334-179A-4
; Sequence 4, Application US/08334179A
; Patent No. 6306622
; GENERAL INFORMATION:
; APPLICANT: ROSENBAUM, JAN S.
; APPLICANT: NOHNO, TSUTOMU
; TITLE OF INVENTION: CDNA ENCODING A BMP TYPE II RECEPTOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE PROCTER AND GAMBLE COMPANY
; STREET: 11810 EAST MIAMI RIVER ROAD
; CITY: ROSS
; STATE: OH
; COUNTRY: US
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.30, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/334,179A
; FILING DATE: 04-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CORSTANJE, BRAHM J.
; REGISTRATION NUMBER: 34,804
; REFERENCE/DOCKET NUMBER: 5473
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 513-627-2858
; TELEFAX: 513-627-0260
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 582 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-334-179A-4
```

```
Query Match 1.1%; Score 8; DB 4; Length 582;
Best Local Similarity 100.0%; Pred.No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 329 ENSLAVRS 336
Db 468 ENSLAVRS 475
```

```
RESULT 4
US-09-252-991A-20231
; Sequence 20231, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107195.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20231
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20231
```

```
Query Match 1.1%; Score 8; DB 4; Length 757;
Best Local Similarity 100.0%; Pred.No. 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 165 LGSRSHEV 172
Db 455 LGSRSHEV 462
```

```
RESULT 5
US-08-334-179A-2
; Sequence 2, Application US/08334179A
; Patent No. 6306622
; GENERAL INFORMATION:
; APPLICANT: ROSENBAUM, JAN S.
; APPLICANT: NOHNO, TSUTOMU
; TITLE OF INVENTION: CDNA ENCODING A BMP TYPE II RECEPTOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE PROCTER AND GAMBLE COMPANY
; STREET: 11810 EAST MIAMI RIVER ROAD
```


CITY: ROSS
STATE: OH
COUNTRY: US
ZIP: 45061
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/334,179A
FILING DATE: 04-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CORSTANJE, BRAHM J.
REGISTRATION NUMBER: 34,804
REFERENCE/DOCKET NUMBER: 5473
TELEPHONE: 513-627-2858
TELEFAX: 513-627-0260
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-334-179A-2

Query Match 1.1%; Score 8; DB 4; Length 1038;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 329 ENSLAVERS 336
Db 468 ENSLAVERS 475

RESULT 6
US-08-334-179A-8
; Sequence 8, Application US/08334179A
; Patent No. 6306622
; GENERAL INFORMATION:
; APPLICANT: ROSENBAUM, JAN S.
; APPLICANT: NOHNO, TSUTOMU
; TITLE OF INVENTION: CDNA ENCODING A BMP TYPE II RECEPTOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE PROCTER AND GAMBLE COMPANY
; STREET: 11810 EAST MIAMI RIVER ROAD
; CITY: ROSS
; STATE: OH
; COUNTRY: US
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.30, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/334,179A
; FILING DATE: 04-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CORSTANJE, BRAHM J.
; REGISTRATION NUMBER: 34,804
; REFERENCE/DOCKET NUMBER: 5473
; TELEPHONE: 513-627-2858
; TELEFAX: 513-627-0260
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1038 amino acids
; TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-334-179A-8

Query Match 1.1%; Score 8; DB 4; Length 1038;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 329 ENSLAVERS 336
Db 468 ENSLAVERS 475

RESULT 7
US-08-844-978-30
; Sequence 30, Application US/08844978
; Patent No. 6075004
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbetts, Scott
; TITLE OF INVENTION: Peptide Compositions Which Induce Immune
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/844,978
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 24769-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: 816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-844-978-30

Query Match 0.9%; Score 7; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 228 SYSLDDL 234
Db 4 SYSLDDL 10

RESULT 8
US-08-844-978-33
; Sequence 33, Application US/08844978
; Patent No. 6075004
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia

APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: Peptide Compositions Which Induce Immune
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 2405 Grand Boulevard, Suite 400
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,978
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 24769-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
TOPOLOGY: circular
MOLECULE TYPE: peptide
US-08-844-978-33

Query Match 0.9%; Score 7; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 SYSDDL 234
|||||
DB 4 SYSDDL 10

RESULT 9
US-08-979-608A-19
Sequence 19, Application US/08979608A
Patent No. 6355451
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
Lees, Robert S.
Law, Simon W.
Atjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,608A
FILING DATE: 28-NOV-1997
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/048,547
FILING DATE: 03-JUN-1997
APPLICATION NUMBER: US 60/031,930
FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Louis
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-979-608A-19

Query Match 0.9%; Score 7; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 753 DEDEDE 759
|||||
DB 7 DEDEDE 13

RESULT 10
US-08-789-078-16
Sequence 16, Application US/08789078
Patent No. 5843985
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/789,078
FILING DATE: 03-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,513
FILING DATE: 19-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: tonsil
US-08-789-078-16

Query Match 0.9%; Score 7; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 SYSDDL 234
Db 3 SYSDDL 9

RESULT 11
US-08-752-633-16
Sequence 16, Application US/08752633
Patent No. 5863889
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
TITLE OF INVENTION: IMMUNE TOLERANCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,633
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: tonsil
US-08-752-633-16

Query Match 0.9%; Score 7; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 SYSDDL 234
Db 3 SYSDDL 9

RESULT 12
US-08-844-978-29
Sequence 29, Application US/08844978
Patent No. 6075004
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: Peptide Compositions Which Induce Immune
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 2405 Grand Boulevard, Suite 400
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,978
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 24769-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-844-978-29

Query Match 0.9%; Score 7; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 SYSDDL 234
Db 3 SYSDDL 9

RESULT 13
PCT-US95-04886-16
Sequence 16, Application PC/TUS9504886
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
TITLE OF INVENTION: IMMUNE TOLERANCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: (816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: tonsil
PCT-US95-04886-16

Query Match 0.9%; Score 7; DB 5; Length 25;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 SYSDDL 234
Db 3 SYSDDL 9

RESULT 14

US-08-979-608A-20
Sequence 20, Application US/08979608A
Patent No. 6355451
GENERAL INFORMATION:

APPLICANT: Lees, Ann M.

Lees, Robert S.

Law, Simon W.

Arjona, Anibal A.

TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN

BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
TREATING ATHEROSCLEROSIS

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,608A

FILING DATE: 26-Nov. 6355451-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/048,547

FILING DATE: 03-JUN-1997

APPLICATION NUMBER: US 60/031,930

FILING DATE: 27-NOV-1996

ATTORNEY/AGENT INFORMATION:

NAME: Myers, Louis

REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-08-979-608A-20

Query Match 0.9%; Score 7; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 753 DEDEEE 759
Db 7 DEDEEE 13

RESULT 15

US-09-252-991A-27697
Sequence 27697, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 27697

LENGTH: 70

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27697

Query Match 0.9%; Score 7; DB 4; Length 70;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 RSASGRK 223
Db 19 RSASGRK 25

Search completed: January 6, 2004, 09:55:05
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:50:50 ; Search time 38 Seconds
(without alignments)
4007.544 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSPFNRQWTSLSLAVTA.....LSVEEQIKRNYDEDEDE 759

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 73937 seqs, 200641211 residues

Word size : 0

Total number of hits satisfying chosen parameters: 73937

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_5/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	657	86	6	755	11 US-09-783-732-4
2	313	41	2	596	11 US-09-783-732-2
3	184	24	2	243	9 US-09-925-297-611
4	41	5	4	52	11 US-09-783-732-5
5	15	2	0	527	9 US-09-880-192-53
6	15	2	0	527	12 US-10-427-348-53
7	14	1	8	127	12 US-10-440-366-12
8	14	1	8	127	12 US-10-104-047-2959
9	14	1	8	158	9 US-09-789-919-42
10	10	1	3	406	12 US-10-161-927-74
11	8	1	1	236	10 US-09-764-864-1047
12	8	1	1	297	12 US-10-094-749-1748
13	8	1	1	398	11 US-09-932-367A-6
14	8	1	1	514	12 US-10-369-493-14097
15	8	1	1	1038	10 US-09-908-500A-2

16	8	1.1	1038	12	US-10-286-152A-42	Sequence 42, Appl
17	8	1.1	1080	9	US-09-904-380-2	Sequence 2, Appl
18	7	0.9	10	11	US-09-572-404B-4124	Sequence 4124, Ap
19	7	0.9	15	9	US-09-962-055-19	Sequence 19, Appl
20	7	0.9	15	10	US-09-976-740-19	Sequence 19, Appl
21	7	0.9	15	14	US-10-023-529-19	Sequence 19, Appl
22	7	0.9	15	14	US-10-023-523-19	Sequence 19, Appl
23	7	0.9	16	15	US-10-225-567A-1216	Sequence 1216, Ap
24	7	0.9	26	9	US-09-962-055-20	Sequence 20, Appl
25	7	0.9	26	10	US-09-976-740-20	Sequence 20, Appl
26	7	0.9	26	14	US-10-023-529-20	Sequence 20, Appl
27	7	0.9	26	14	US-10-023-523-20	Sequence 20, Appl
28	7	0.9	40	11	US-09-764-891-4557	Sequence 4557, Ap
29	7	0.9	53	11	US-09-764-891-4689	Sequence 4689, Ap
30	7	0.9	68	9	US-09-864-761-39951	Sequence 39951, A
31	7	0.9	68	12	US-10-029-386-28019	Sequence 28019, A
32	7	0.9	76	11	US-09-374-046A-100	Sequence 100, App
33	7	0.9	98	15	US-10-106-698-5523	Sequence 5523, Ap
34	7	0.9	154	12	US-10-369-493-11211	Sequence 11211, A
35	7	0.9	161	12	US-10-238-075-1252	Sequence 1252, Ap
36	7	0.9	181	15	US-10-156-761-10026	Sequence 10026, A
37	7	0.9	186	14	US-10-108-605-33	Sequence 33, Appl
38	7	0.9	196	11	US-09-934-455-342	Sequence 342, App
39	7	0.9	205	12	US-10-218-137-3	Sequence 3, Appl
40	7	0.9	205	15	US-10-233-448-3	Sequence 3, Appl
41	7	0.9	217	9	US-09-962-055-7	Sequence 7, Appl
42	7	0.9	217	10	US-09-976-740-7	Sequence 7, Appl
43	7	0.9	217	14	US-10-023-529-7	Sequence 7, Appl
44	7	0.9	217	14	US-10-023-523-7	Sequence 7, Appl
45	7	0.9	232	12	US-10-369-493-13799	Sequence 13799, A

ALIGNMENTS

RESULT 1

US-09-783-732-4
; Sequence 4, Application US/09783732
; Publication No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; TITLE OF INVENTION: (EPLIN)
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09/783.732
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homosapien
US-09-783-732-4

Query Match	86.6%	Score 657	DB 11	Length 755
Best Local Similarity	100.0%	Pred. No. 0		
Matches 657	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MESSPFNRQWTSLSLRVTAKELSLVNKNKSSAIVIFSKYQKAAETNMKRSNTENL	60	
Db	1	MESSPFNRQWTSLSLRVTAKELSLVNKNKSSAIVIFSKYQKAAETNMKRSNTENL	60	
QY	61	SOHFRKGTITLVKKKNWPCLGAEHTDSLRNSTEIRHRAHPPAEVTTSHAASGAKAQ	120	
Db	61	SOHFRKGTITLVKKKNWPCLGAEHTDSLRNSTEIRHRAHPPAEVTTSHAASGAKAQ	120	
QY	121	EQIHPRSRLRSPPEALVQGRYPHIKOGEDLKHSTESKKNKCLGESRHEVEKSEISEN	180	
Db	121	EQIHPRSRLRSPPEALVQGRYPHIKOGEDLKHSTESKKNKCLGESRHEVEKSEISEN	180	

QY 181 TDASGKIEKYNVPLNRLKMFGEKGTQTKILRAQSRASGRKISSENSYSLDLEIGPQ 240
Db 181 TDASGKIEKYNVPLNRLKMFGEKGTQTKILRAQSRASGRKISSENSYSLDLEIGPQ 240
QY 241 LSSSTFDSKSEKSRRLNLELPRLSETSIKORMAKYQAAVSKQSSSTNYTNELKASGGEIKI 300
Db 241 LSSSTFDSKSEKSRRLNLELPRLSETSIKORMAKYQAAVSKQSSSTNYTNELKASGGEIKI 300
QY 301 HKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVSEVQVHPKP 360
Db 301 HKMEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVSEVQVHPKP 360
QY 361 LSPDSRASSLSSESPPKAMKKTQAPARETCVCOQTVYPMERLLANQQVFHISCFRCSCYC 420
Db 361 LSPDSRASSLSSESPPKAMKKTQAPARETCVCOQTVYPMERLLANQQVFHISCFRCSCYC 420
QY 421 NNKLSLGTVASLHGRYCKPHFNOLFKSKGNVDEGFGHRPHKDLWASKNENEIILERPQA 480
Db 421 NNKLSLGTVASLHGRYCKPHFNOLFKSKGNVDEGFGHRPHKDLWASKNENEIILERPQA 480
QY 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKRLRIAWPPPTELGS 540
Db 481 LANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKRLRIAWPPPTELGS 540
QY 541 SGSALEEGIKMSKPKWPPDEISKEPVEDVDLDLKLRRSSSLKERSRPTVVAASFOT 600
Db 541 SGSALEEGIKMSKPKWPPDEISKEPVEDVDLDLKLRRSSSLKERSRPTVVAASFOT 600
QY 601 SVKSPKTVSPPIRKGMSSESESVGGRVAERKOVENAKSKNGNVGKTTWQNK 657
Db 601 SVKSPKTVSPPIRKGMSSESESVGGRVAERKOVENAKSKNGNVGKTTWQNK 657

RESULT 2

US-09-783-732-2
; Sequence 2, Application US/09783732
; Patent No. US20030054417A1
; GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; TITLE OF INVENTION: (EPLIN)
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09/783,732
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 09/658,400
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homosapien
US-09-783-732-2

Query Match 41.2%; Score 313; DB 11; Length 596;
Best Local Similarity 100.0%; Pred. No. 9.8e-290;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 345 DSQVSEVQVHPKPLSPDSRASSLSSESPPKAMKKTQAPARETCVCOQTVYPMERLL 404
Db 186 DSQVSEVQVHPKPLSPDSRASSLSSESPPKAMKKTQAPARETCVCOQTVYPMERLL 245
QY 405 ANQOVHISCFRCSCYCNKLSLGTASLHGRYCKPHFNOLFKSKGNVDEGFGHRPHKDL 464
Db 246 ANQOVHISCFRCSCYCNKLSLGTASLHGRYCKPHFNOLFKSKGNVDEGFGHRPHKDL 305
QY 465 WASKNENEIILERPQAANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAE 524
Db 306 WASKNENEIILERPQAANARETPHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAE 365

QY 525 TKKURIAWPPPTTELGSSGSALEEGIKMSKPKWPPDEISKEPVEDVDLDLKLRRSSSL 584
Db 366 TKKURIAWPPPTTELGSSGSALEEGIKMSKPKWPPDEISKEPVEDVDLDLKLRRSSSL 425
QY 585 KERSRPTVVAASFOTSVKSPKTVSPPIRKGMSSESESVGGRVAERKOVENAKASKK 644
Db 426 KERSRPTVVAASFOTSVKSPKTVSPPIRKGMSSESESVGGRVAERKOVENAKASKK 485
QY 645 NGNVGKTTWQNK 657
Db 486 NGNVGKTTWQNK 498

RESULT 3

US-09-925-297-611
; Sequence 611, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 611
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (185)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (237)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (238)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (243)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-611

Query Match 24.2%; Score 184; DB 9; Length 243;
Best Local Similarity 100.0%; Pred. No. 5.7e-167;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 EGEKISANENSLAVRSTPAEDDSRDSQVSEVQVHPKPSPDSRASSLSSESPPKAMK 380
Db 1 EGEKISANENSLAVRSTPAEDDSRDSQVSEVQVHPKPSPDSRASSLSSESPPKAMK 60
QY 381 KFOAPARETCVCOQTVYPMERLLANQQVFHISCFRCSCYCNKLSLGTASLHGRYCKP 440
Db 61 KFOAPARETCVCOQTVYPMERLLANQQVFHISCFRCSCYCNKLSLGTASLHGRYCKP 120
QY 441 HFNQLFKSKGNVDEGFGHRPHKDLWASKNENEIILERPQAANARETPHSPGVEDAPIAK 500
Db 121 HFNQLFKSKGNVDEGFGHRPHKDLWASKNENEIILERPQAANARETPHSPGVEDAPIAK 180
QY 501 VGVVL 504
Db 181 VGVVL 184

RESULT 4

US-09-783-732-5
; Sequence 5, Application US/09783732
; Publication No. US20030054417A1

GENERAL INFORMATION:
; APPLICANT: Chang, David D.
; APPLICANT: Maul, Raymond S.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: EPITHELIAL PROTEIN LOST IN NEOPLASM
; FILE REFERENCE: 10809/003001
; CURRENT APPLICATION NUMBER: US/09/783,732
; PRIOR FILING DATE: 2001-02-13
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (0)...(0)
; OTHER INFORMATION: mutant sterol regulatory element binding protein 2
; US-09-783-732-5

Query Match 5.4%; Score 41; DB 11; Length 52;
Best Local Similarity 100.0%; Pred. No. 4.3e-31;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 CQKTVPMERLLANQOVFHISCFRCSCYCNKLSLGTYSLSH 433
DB 4 CQKTVPMERLLANQOVFHISCFRCSCYCNKLSLGTYSLSH 44

RESULT 5
US-09-880-192-53
; Sequence 53, Application US/09880192
; Patent No. US20020077470A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/09/880,192
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 53
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020077470A1 058201CD1
US-09-880-192-53

Query Match 2.0%; Score 15; DB 9; Length 527;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFGH 458
DB 45 QLFKSKGNYDEGFGH 59

RESULT 6
US-10-427-348-53
; Sequence 53, Application US/10427348
; Publication No. US20030175795A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.

; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-2 CON
; CURRENT APPLICATION NUMBER: US/10/427,348
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/880,192
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: US 09/299,708
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 53
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175795A1 058201CD1
US-10-427-348-53

Query Match 2.0%; Score 15; DB 12; Length 527;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFGH 458
DB 45 QLFKSKGNYDEGFGH 59

RESULT 7
US-10-440-366-12
; Sequence 12, Application US/10440366
; Publication No. US20030181706A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: INFLAMMATION-ASSOCIATED POLYNUCLEOTIDES
; FILE REFERENCE: PB-0006-1 CIP
; CURRENT APPLICATION NUMBER: US/10/440,366
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: US 09/855,323
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/195,292
; PRIOR FILING DATE: 1998-11-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PERL Program
; SEQ ID NO 12
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 402234CD1
US-10-440-366-12

Query Match 1.8%; Score 14; DB 12; Length 127;
Best Local Similarity 100.0%; Pred. No. 5.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFGH 457
DB 94 QLFKSKGNYDEGFGH 107

RESULT 8
US-10-104-047-2959
; Sequence 2959, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047

; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2959
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2959

Query Match 1.8%; Score 14; DB 12; Length 127;
Best Local Similarity 100.0%; Pred. No. 5.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFG 457
Db 94 QLFKSKGNYDEGFG 107

RESULT 9
US-09-789-919-42
; Sequence 42, Application US/09789919
; Patent No. US20020064855A1
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor
; APPLICANT: Moore, Kaceri
; TITLE OF INVENTION: GENES THAT REGULATE HEMATOPOIETIC BLOOD FORMING STEM
; FILE REFERENCE: 2275-1-005
; CURRENT APPLICATION NUMBER: US/09/789,919
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-789-919-42

Query Match 1.8%; Score 14; DB 9; Length 158;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFG 457
Db 95 QLFKSKGNYDEGFG 108

RESULT 10
US-10-161-927-74
; Sequence 74, Application US/10161927
; Publication No. US2003023821A1
; GENERAL INFORMATION:
; APPLICANT: Zernhusen, Bryan D.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Miller, Charles E.
; APPLICANT: Hjalte, Tord
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Vernet, Corine
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Li, Li
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Gorman, Linda
; APPLICANT: Anderson, David W.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Patturajan, Meera
; APPLICANT: Stone, David J.

; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 21402-377 D (Cura 677 Other)
; CURRENT APPLICATION NUMBER: US/10/161,927
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: 60/295,661
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/295,607
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/296,404
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/296,418
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: 60/296,575
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/297,414
; PRIOR FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: 60/297,567
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/298,528
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/325,685
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/299,133
; PRIOR FILING DATE: 2001-06-18
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 190
; SEQ ID NO 74
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-927-74

Query Match 1.3%; Score 10; DB 12; Length 406;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 KMSKPKWPPE 559
Db 62 KMSKPKWPPE 71

RESULT 11
US-09-764-864-1047
; Sequence 1047, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1047
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (43)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (44)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (226)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE

; LOCATION: (228)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1047

Query Match 1.1%; Score 8; DB 10; Length 236;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 YDEDEDEE 759
| | | | | | | |
DB 107 YDEDEDEE 114

RESULT 12

US-10-094-749-1748
; Sequence 1748, Application US/10094749
; Publication No. US20030219741A1

; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: HIO, YURI

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHIKO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOKYUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA

; FILE REFERENCE: 084335/0160

; CURRENT APPLICATION NUMBER: US/10/094,749

; CURRENT FILING DATE: 2002-03-12

; PRIOR FILING DATE: 2002-03-12

; PRIOR FILING DATE: 2002-01-24

; PRIOR APPLICATION NUMBER: JP 2001-328381

; PRIOR FILING DATE: 2001-09-14

; NUMBER OF SEQ ID NOS: 3381

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1748

; LENGTH: 297

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-094-749-1748

Query Match 1.1%; Score 8; DB 12; Length 297;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 YDEDEDEE 759
| | | | | | | |
DB 157 YDEDEDEE 164

RESULT 13

US-09-932-367A-6

; Sequence 6, Application US/09932367A

; Publication No. US20030027152A1

; GENERAL INFORMATION:

; APPLICANT: RHODES, Simon J.

; APPLICANT: BRIDWELL, Jeanne L.

; APPLICANT: MEIER, Bradley C.

; APPLICANT: PARKER, Gretchen E.

; APPLICANT: PRICE, Jeffrey R.

; APPLICANT: SHOWALTER, Aaron D.

; APPLICANT: SLOOP, Kyle W.

; TITLE OF INVENTION: GENERATION OF DIAGNOSTIC TOOLS TO ASSAY THE HUMAN

; TITLE OF INVENTION: LHX3P-LIM/LIM-3 FACTOR

; FILE REFERENCE: 053884-5003

; CURRENT APPLICATION NUMBER: US/09/932,367A

; CURRENT FILING DATE: 2001-08-17

; PRIOR FILING DATE: 2000-02-22

; PRIOR FILING DATE: 2000-02-22

; PRIOR APPLICATION NUMBER: US 60/121,110

; PRIOR FILING DATE: 1999-02-22

; NUMBER OF SEQ ID NOS: 113

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 398

; TYPE: PRT

; ORGANISM: Danio rerio

US-09-932-367A-6

Query Match 1.1%; Score 8; DB 11; Length 398;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 368 SSLSESSP 375
| | | | | | | |
DB 272 SSLSESSP 279

RESULT 14

US-10-369-493-14097

; Sequence 14097, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 14097

; LENGTH: 514

; TYPE: PRT

; ORGANISM: Pseudomonas fluorescens

US-10-369-493-14097

Query Match 1.1%; Score 8; DB 12; Length 514;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 SVKSPKTV 608
| | | | | | | |
DB 487 SVKSPKTV 494

RESULT 15

US-09-908-500A-2

; Sequence 2, Application US/09908500A

; Patent No. US20020102576A1

; GENERAL INFORMATION:

; APPLICANT: James Loyd

; APPLICANT: Kirk B. Lane

; APPLICANT: John A. Phillips, III

; TITLE OF INVENTION: METHOD OF DIAGNOSING PULMONARY

; FILE REFERENCE: 22000.0108U3

; CURRENT APPLICATION NUMBER: US/09/908,500A

; CURRENT FILING DATE: 2001-07-17

; PRIOR APPLICATION NUMBER: 60/215,740

```
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/220,133
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1038
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-908-500A-2

Query Match      1.1%  Score 8;  DB 10;  Length 1038;
Best Local Similarity 100.0%;  Pred. No. 1.9e+02;
Matches 8;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      329 ENSLAVRS 336
      |||||
Db      468 ENSLAVRS 475

Search completed: January 6, 2004, 09:54:30
Job time : 38 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:48:00 ; Search time 48 Seconds
(without alignments)
2509.864 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSFNRRQWTSLSRVTA.....LSVEEQIKRNYDEDEEE 759

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A. Geneseq 19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	759	100.0	759	21	AA197286 Lipid associated p
2	759	100.0	759	22	AA197286 Human transcriptio
3	759	100.0	759	22	AA197286 Human polypeptide
4	678	89.3	759	22	AA197286 Human EPLIN (epith
5	538	70.9	760	21	AA197286 Human ORFX ORF2698
6	457	60.2	457	22	AA197286 Human breast sequ
7	370	48.7	471	20	AA197286 Human breast tumou
8	334	44.0	600	22	AA197286 Human EPLIN (epith
9	299	39.4	299	21	AA197286 Human secreted pro

10	240	31.6	301	23	AB556420 Human cancer suppr
11	238	31.4	457	23	AA10979 Human steroid regul
12	227	29.9	232	21	AA551883 Human secreted pro
13	184	24.2	243	21	AA551883 Human pancreatic c
14	170	22.4	235	20	AA173916 Human prostate tum
15	155	20.4	337	22	ABG24090 Novel human diagno
16	131	17.3	476	22	ABG08559 Novel human diagno
17	87	11.5	152	22	AA40799 Human polypeptide
18	67	8.8	88	21	AA40799 Human secreted pro
19	44	5.8	331	22	ABG08554 Novel human diagno
20	28	3.7	234	21	AA551882 Gene 5 human secre
21	28	3.7	299	21	AA551882 Gene 41 human secr
22	15	2.0	527	23	AB43758 Human breast tumou
23	15	2.0	527	23	AA25978 Human protein #5 c
24	15	2.0	547	22	AA20575 Human secreted pro
25	15	2.0	547	22	AA20632 Human secreted pro
26	15	2.0	547	23	AA16626 Human 41441 protei
27	14	1.8	111	22	ABG12835 Novel human diagno
28	14	1.8	127	21	AA19447 Human inflammation
29	14	1.8	135	22	ABG12836 Novel human diagno
30	14	1.8	158	21	AA179176 Haematopoietic ste
31	14	1.8	173	23	ABG06053 Human NS protein s
32	14	1.8	173	23	ABG06054 Human NS protein s
33	14	1.8	173	23	ABG06055 Human NS protein s
34	14	1.8	173	23	ABG06056 Human NS protein s
35	14	1.8	173	23	ABG06057 Human NS protein s
36	14	1.8	173	23	ABG06058 Human NS protein s
37	14	1.8	217	22	AA20850 Human secreted pro
38	13	1.7	51	21	AA603479 Human secreted pro
39	11	1.4	90	21	AA32625 Eucalyptus grandis
40	11	1.4	104	21	AA32623 Eucalyptus grandis
41	11	1.4	125	23	AB41582 Human ovarian anti
42	10	1.3	171	22	ABG04347 Novel human diagno
43	9	1.2	9	24	ABJ20052 MHC binding peptid
44	9	1.2	283	23	AA15050 Human 187-2 protei
45	8	1.1	115	23	ABP31959 Human ORF932 prote

ALIGNMENTS

RESULT 1

AA197286

AA197286 standard; Protein; 759 AA.

AA197286;

03-JAN-2001 (first entry)

Lipid associated protein (LIPAP) 2766980CD1.

Lipid associated protein; LIPAP; treatment; prophylaxis; agonist;

antagonist; antibody; cardiovascular disease; neurological disease;

gastrointestinal disease; lipid metabolism; detection;

amplification; monitoring; hybridisation; antisense; triplex;

ribosome; screening; immunoassay.

Homo sapiens.

Key

Modified-site

Location/Qualifiers

15

/note= "Potential phosphorylation site"

19

/note= "Potential phosphorylation site"

29

/note= "Potential glycosylation site"

48

/note= "Potential phosphorylation site"

55

/note= "Potential phosphorylation site"

59

/note= "Potential glycosylation site"

68

FT	Modified-site	/note= "Potential phosphorylation site"	89
FT	Modified-site	/note= "Potential phosphorylation site"	85
FT	Modified-site	/note= "Potential phosphorylation site"	92
FT	Modified-site	/note= "Potential glycosylation site"	93
FT	Modified-site	/note= "Potential phosphorylation site"	132
FT	Modified-site	/note= "Potential phosphorylation site"	158
FT	Modified-site	/note= "Potential phosphorylation site"	168
FT	Modified-site	/note= "Potential phosphorylation site"	184
FT	Modified-site	/note= "Potential phosphorylation site"	220
FT	Modified-site	/note= "Potential phosphorylation site"	225
FT	Modified-site	/note= "Potential phosphorylation site"	229
FT	Modified-site	/note= "Potential phosphorylation site"	230
FT	Modified-site	/note= "Potential phosphorylation site"	244
FT	Modified-site	/note= "Potential phosphorylation site"	248
FT	Modified-site	/note= "Potential phosphorylation site"	251
FT	Modified-site	/note= "Potential glycosylation site"	253
FT	Modified-site	/note= "Potential phosphorylation site"	266
FT	Modified-site	/note= "Potential phosphorylation site"	286
FT	Modified-site	/note= "Potential glycosylation site"	294
FT	Modified-site	/note= "Potential phosphorylation site"	318
FT	Modified-site	/note= "Potential phosphorylation site"	326
FT	Modified-site	/note= "Potential phosphorylation site"	337
FT	Domain	/note= "Potential phosphorylation site"	344..444
FT	Modified-site	/label= LIM domain	369
FT	Modified-site	/note= "Potential phosphorylation site"	389
FT	Modified-site	/note= "Potential phosphorylation site"	467
FT	Modified-site	/note= "Potential phosphorylation site"	514
FT	Modified-site	/note= "Potential phosphorylation site"	525
FT	Modified-site	/note= "Potential phosphorylation site"	543
FT	Modified-site	/note= "Potential phosphorylation site"	563
FT	Modified-site	/note= "Potential phosphorylation site"	582
FT	Modified-site	/note= "Potential phosphorylation site"	583
FT	Modified-site	/note= "Potential phosphorylation site"	601
FT	Modified-site	/note= "Potential phosphorylation site"	604
FT	Modified-site	/note= "Potential phosphorylation site"	617
FT	Modified-site	/note= "Potential phosphorylation site"	642
FT	Modified-site	/note= "Potential phosphorylation site"	

FT	Modified-site	658	/note= "Potential phosphorylation site"
FT	Modified-site	662	/note= "Potential phosphorylation site"
FT	Modified-site	686	/note= "Potential phosphorylation site"
FT	Modified-site	698	/note= "Potential phosphorylation site"
FT	Modified-site	706	/note= "Potential phosphorylation site"
FT	Modified-site	709	/note= "Potential glycosylation site"
FT	Modified-site	714	/note= "Potential phosphorylation site"
FT	Modified-site	741	/note= "Potential phosphorylation site"
FT	Modified-site		/note= "Potential phosphorylation site"
PN	WO2000049043-A2.		
XX			
PD	24-AUG-2000.		
XX			
PF	18-FEB-2000; 2000WO-US04160.		
XX			
PR	19-FEB-1999; 99US-0120703.		
PR	08-JUL-1999; 99US-0142762.		
XX			
PA	(INCY-) INCYTE PHARM INC.		
PI	Tang YT, Hillman JL, Yue H, Azimzai Y, Baughn MR, Tran B;		
XX			
DR	WPI; 2000-549264/50.		
DR	N-PSDB; AAA53826.		
XX			
PT	New human lipid-associated proteins, nucleic acids, and antibodies,		
PT	useful for diagnosis, treatment and prevention of e.g. cardiovascular		
PT	disease		
PS	Claim 1; Page 76-77; 93pp; English.		
XX			
CC	Lipid-associated proteins (LIPAP) can be used for treating or		
CC	preventing disorders associated with decreased expression of LIPAP,		
CC	for screening for agonists or antagonists of LIPAP, and to raise		
CC	specific antibodies. Antagonists and antagonists of LIPAP are useful		
CC	for treating diseases associated with reduced or increased levels of		
CC	LIPAP, e.g. cardiovascular, neurological and gastrointestinal		
CC	diseases and disorders of lipid metabolism. Fragments of the nucleic		
CC	acid encoding LIPAP are useful for detection of full length coding		
CC	sequences, in hybridization and/or amplification assays or for		
CC	diagnosis or monitoring. Nucleotides encoding LIPAP are used		
CC	to screen for compounds that specifically modify LIPAP expression,		
CC	for recombinant production of LIPAP, in gene therapy, as a source of		
CC	therapeutic antisense, triplex-forming, or ribozyme agents and for		
CC	genomic mapping. Antibodies to the proteins are used for diagnosis		
CC	and monitoring of LIPAP-associated disease by immunoassay, as		
CC	antagonists, in competitive drug screens and for affinity		
CC	purification of natural LIPAP.		

AA	Sequence	759 AA;
----	----------	---------

Query Match	100.0%;	Score 759;	DB 21;	Length 759;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 759; Conservative	0;	Mismatches	0;	Indels

	Matches	759; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
Qy	1	MESSPNNRRQWTSLSLRVTAKELSLVNKYNKSSAIVEIFSKYQYQAAAEETNMEKKRSNTENL	60			
Db	1	MESSPNNRRQWTSLSLRVTAKELSLVNKYNKSSAIVEIFSKYQYQAAAEETNMEKKRSNTENL	60			
Qy	61	SQHFRGIGTLTVLKKKWNENPGLGAESHTDSLNRNSSTETIRHRADHPPAEVTSHAASGAKAQD	120			
Db	61	SQHFRGIGTLTVLKKKWNENPGLGAESHTDSLNRNSSTETIRHRADHPPAEVTSHAASGAKAQD	120			
Qy	121	EQCIHPRSRFLRPPPEALVQGRTPHIKDGEDLKDHSSTCKMKNCLGESRHEVEKSFISEN	180			

Db 121 EQIHPRRLRSPPEALVQGRYPHIKDGEDLKDHSTESKMKWENCLGESRHEVEKSEISEN 180
QY 181 TDASGKTEKNVPLNRLKMMFEKGEPTQTKILRAQSRASGRKISENSYSLDLEIGPGQ 240
Db 181 TDASGKTEKNVPLNRLKMMFEKGEPTQTKILRAQSRASGRKISENSYSLDLEIGPGQ 240
QY 241 LSSSTFDSEKNESRRNLELPRLSSETSICKRMAYQAAVSKQSSSTNTYNELKASGGGEIKI 300
Db 241 LSSSTFDSEKNESRRNLELPRLSSETSICKRMAYQAAVSKQSSSTNTYNELKASGGGEIKI 300
QY 301 HKMQEKNVPPGPEVCITHOGEKISANENSLAVRSTPAEDDSDRSQVKSEVQPVHPKP 360
Db 301 HKMQEKNVPPGPEVCITHOGEKISANENSLAVRSTPAEDDSDRSQVKSEVQPVHPKP 360
QY 361 LSPDSRASSLSSEPPKAMKFKQAPARETCVECKTYPMERLLANQQVPHISCFCRSCYC 420
Db 361 LSPDSRASSLSSEPPKAMKFKQAPARETCVECKTYPMERLLANQQVPHISCFCRSCYC 420
QY 421 NNKLSLGTYSASLHGRIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEILERPAP 480
Db 421 NNKLSLGTYSASLHGRIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEILERPAP 480
QY 481 LANARETPHSPGVEDAPIAKGVVLAASMEAKASSQOEKEDKPAETKKLR.IAWPPPTLGS 540
Db 481 LANARETPHSPGVEDAPIAKGVVLAASMEAKASSQOEKEDKPAETKKLR.IAWPPPTLGS 540
QY 541 SGSALEEGIKMSKPKWPPEDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFQST 600
Db 541 SGSALEEGIKMSKPKWPPEDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFQST 600
QY 601 SVKSPKTVSPPIRGWMSQSESVGGRVAERKQVENAKSKNGVNGKTTWQNKESKG 660
Db 601 SVKSPKTVSPPIRGWMSQSESVGGRVAERKQVENAKSKNGVNGKTTWQNKESKG 660
QY 661 ETGKRSKEGHSLEMEENLVENGADSDSDNSFLKQSPQEPKSLNWSSFDVNTFAEEFT 720
Db 661 ETGKRSKEGHSLEMEENLVENGADSDSDNSFLKQSPQEPKSLNWSSFDVNTFAEEFT 720
QY 721 TQNKQSDVELWEGEVVKELSVBEOIKRNYRYDEDEEE 759
Db 721 TQNKQSDVELWEGEVVKELSVBEOIKRNYRYDEDEEE 759

RESULT 2

ABU52869

ID ABU52869 standard; Protein; 759 AA.

XX AC ABU52869;

XX AC ABU52869; (first entry)

XX DT 14-APR-2003

XX DE Human transcription factor from DKF2phutet1_18c19.

XX KW Human; gene therapy; vaccine; disease treatment; detection.

XX OS Homo sapiens.

XX OS Homo sapiens.

XX PN WC200112659-A2.

XX PD 22-FEB-2001.

XX PF 18-AUG-2000; 2000MO-IB01496.

XX PR 18-AUG-1999; 99US-0149499.

XX PR 28-SEP-1999; 99US-0156503.

XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.

XX PI Wiemann S;

XX PI Wiemann S;

XX DR WPI: 2001-327840/34.

XX DR N-PSDB; ABX71307.

XX

PT Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies -
XX Claim 21; Page 450-451; 1095pp; English.
XX This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a
CC polypeptide described in the disclosure of the invention.
XX
SQ Sequence 759 AA;

Query Match 100.0%; Score 759; DB 22; Length 759;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESSPFNRQWTSLSLRVTAKELSLVNNKSSAIVEIFSKYQKAAEETNMEKRSNTENL 60
Db 1 MESSPFNRQWTSLSLRVTAKELSLVNNKSSAIVEIFSKYQKAAEETNMEKRSNTENL 60
QY 61 SQHFRKGTILTVLKKWENPGLCAESHTDLSRNSSTEIRHADHPPEAETVSHAASAKAQDQ 120
Db 61 SQHFRKGTILTVLKKWENPGLCAESHTDLSRNSSTEIRHADHPPEAETVSHAASAKAQDQ 120
QY 121 EQIHPRRLRSPPEALVQGRYPHIKDGEDLKDHSTESKMKWENCLGESRHEVEKSEISEN 180
Db 121 EQIHPRRLRSPPEALVQGRYPHIKDGEDLKDHSTESKMKWENCLGESRHEVEKSEISEN 180
QY 181 TDASGKTEKNVPLNRLKMMFEKGEPTQTKILRAQSRASGRKISENSYSLDLEIGPGQ 240
Db 181 TDASGKTEKNVPLNRLKMMFEKGEPTQTKILRAQSRASGRKISENSYSLDLEIGPGQ 240
QY 241 LSSSTFDSEKNESRRNLELPRLSSETSICKRMAYQAAVSKQSSSTNTYNELKASGGGEIKI 300
Db 241 LSSSTFDSEKNESRRNLELPRLSSETSICKRMAYQAAVSKQSSSTNTYNELKASGGGEIKI 300
QY 301 HKMQEKNVPPGPEVCITHOGEKISANENSLAVRSTPAEDDSDRSQVKSEVQPVHPKP 360
Db 301 HKMQEKNVPPGPEVCITHOGEKISANENSLAVRSTPAEDDSDRSQVKSEVQPVHPKP 360
QY 361 LSPDSRASSLSSEPPKAMKFKQAPARETCVECKTYPMERLLANQQVPHISCFCRSCYC 420
Db 361 LSPDSRASSLSSEPPKAMKFKQAPARETCVECKTYPMERLLANQQVPHISCFCRSCYC 420
QY 421 NNKLSLGTYSASLHGRIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEILERPAP 480
Db 421 NNKLSLGTYSASLHGRIYCKPHFNQLFKSKGNYDEGFGHRPHKDLWASKNENEILERPAP 480
QY 481 LANARETPHSPGVEDAPIAKGVVLAASMEAKASSQOEKEDKPAETKKLR.IAWPPPTLGS 540
Db 481 LANARETPHSPGVEDAPIAKGVVLAASMEAKASSQOEKEDKPAETKKLR.IAWPPPTLGS 540
QY 541 SGSALEEGIKMSKPKWPPEDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFQST 600
Db 541 SGSALEEGIKMSKPKWPPEDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFQST 600
QY 601 SVKSPKTVSPPIRGWMSQSESVGGRVAERKQVENAKSKNGVNGKTTWQNKESKG 660
Db 601 SVKSPKTVSPPIRGWMSQSESVGGRVAERKQVENAKSKNGVNGKTTWQNKESKG 660
QY 661 ETGKRSKEGHSLEMEENLVENGADSDSDNSFLKQSPQEPKSLNWSSFDVNTFAEEFT 720
Db 661 ETGKRSKEGHSLEMEENLVENGADSDSDNSFLKQSPQEPKSLNWSSFDVNTFAEEFT 720
QY 721 TQNKQSDVELWEGEVVKELSVBEOIKRNYRYDEDEEE 759
Db 721 TQNKQSDVELWEGEVVKELSVBEOIKRNYRYDEDEEE 759

RESULT 3	
AAM39013	
ID	AAM39013 standard; Protein; 759 AA.
XX	
AC	AAM39013;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polypeptide SEQ ID NO 2158.
XX	
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia.
XX	
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	
DR	WPI; 2001-442253/47.
XX	
DR	N-PSDB; AA158169.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
XX	
PS	Example 4; SEQ ID NO 2158; 10078pp; English.
XX	
CC	The invention relates to human nucleic acids (AA157798-AA161369) and
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilization of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 759 AA;
XX	
Query Match	100.0%; Score 759; DB 22; Length 759;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 759; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MESSPFNRROWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYKAAEETNMEKRSNTENL	60
Db	1	MESSPFNRROWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYKAAEETNMEKRSNTENL	60
QY	61	SOHFRKGTTLVKKWENPOLGAESHTSLRNSSTIRHRADHPPEVTSAAAGAKAQ	120
Db	61	SOHFRKGTTLVKKWENPOLGAESHTSLRNSSTIRHRADHPPEVTSAAAGAKAQ	120
QY	121	EQIHRSLRSPPEALVOGRYPHIKDGEDLKDHSTESKXENCLGESRHEVEKSIEN	180
Db	121	EQIHRSLRSPPEALVOGRYPHIKDGEDLKDHSTESKXENCLGESRHEVEKSIEN	180
QY	181	TDASGKIEKNVPLNRLKWMFEKGEPTQTKILRAQSRASGRKISSENSYSLDDLEIGPGQ	240
Db	181	TDASGKIEKNVPLNRLKWMFEKGEPTQTKILRAQSRASGRKISSENSYSLDDLEIGPGQ	240
QY	241	LSSTFDSEKNESRRNLELPRISSETIKORMAKYQAAVSKQSSTNTYNELKASGGEIKI	300
Db	241	LSSTFDSEKNESRRNLELPRISSETIKORMAKYQAAVSKQSSTNTYNELKASGGEIKI	300
QY	301	HKWEQENVPVPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDQVSEVOQPVHPKP	360
Db	301	HKWEQENVPVPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDQVSEVOQPVHPKP	360
QY	361	LSPDSRASSLSSESSPPKAMKKFOAPARETCVECKTVYPMERLLANQOVFHSCEPSCYC	420
Db	361	LSPDSRASSLSSESSPPKAMKKFOAPARETCVECKTVYPMERLLANQOVFHSCEPSCYC	420
QY	421	NNKLSLGTVASLHGRYICKPHFNQFKSKGNVDEGFGRPHKDLWASKNENEIEILRPAQ	480
Db	421	NNKLSLGTVASLHGRYICKPHFNQFKSKGNVDEGFGRPHKDLWASKNENEIEILRPAQ	480
QY	481	LANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAMPPPTLGS	540
Db	481	LANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAMPPPTLGS	540
QY	541	SGSALBEGIKMSKPKWPPDEISKPEVPEDVDLDLKLRRSSSLKERSPFTVAASFQST	600
Db	541	SGSALBEGIKMSKPKWPPDEISKPEVPEDVDLDLKLRRSSSLKERSPFTVAASFQST	600
QY	601	SVKSPXTVSPPIRKGMSESESVGGVVAERKOVENAKASKKGNVGTWQNKESKG	660
Db	601	SVKSPXTVSPPIRKGMSESESVGGVVAERKOVENAKASKKGNVGTWQNKESKG	660
QY	661	ETGKRSKEGHSLEMNENLVENGADSDDDNDFLQKQSPQEPKSLNWSFVDTTAAEFT	720
Db	661	ETGKRSKEGHSLEMNENLVENGADSDDDNDFLQKQSPQEPKSLNWSFVDTTAAEFT	720
QY	721	TONQKSQDVELWEGEVVKLSVEEQIKRNRYYDEDEDEE	759
Db	721	TONQKSQDVELWEGEVVKLSVEEQIKRNRYYDEDEDEE	759
RESULT 4			
AAB67701			
ID	AAB67701 standard; Protein; 759 AA.		
XX			
AC	AAB67701;		
XX			
DT	11-JUN-2001 (first entry)		
XX			
DE	Human EPLIN (epithelial protein lost in neoplasm).beta isoform.		
XX			
KW	Human; EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;		
KW	EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;		
KW	gene therapy; cancer.		
OS	Homo sapiens.		
XX			
Key	Location/Qualifiers		
FT	Misc-difference 278		
FT	/note= "val encoded by GCG"		
FT	Misc-difference 344		

FT Misc-difference /note= "Arg encoded by CCAGGT"
 FT 495 /note= "Asp encoded by AAT"
 FT Misc-difference 511 /note= "Lys encoded by AGG"
 FT 622 /note= "Ser encoded by AAT"
 FT Misc-difference 658 /note= "Ser encoded by TTT"
 FT 679 /note= "Ser encoded by TTT"
 FT 698 /note= "Ile encoded by TTT"
 FT Misc-difference 698 /note= "Ser encoded by TTT"
 FT 704 /note= "Ser encoded by TTT"
 FT 18019-A1.
 PN WO200118019-A1.
 XX 15-MAR-2001.
 XX 08-SEP-2000; 2000WO-US24689.
 XX 08-SEP-1999; 99US-0153024.
 XX (REGC) UNIV CALIFORNIA.
 XX Chang DD, Maul RS;
 XX WPI; 2001-244555/25.
 DR N-PSDB; AAF55697.
 XX New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
 PT prognostic and therapeutic applications over the course of cell
 PT proliferative disorders associated with EPLIN -
 XX Claim 1; Page 44; 59pp; English.

The present sequence represents a human EPLIN (epithelial protein lost in neoplasm)-beta isoform. The specification also describes EPLIN-alpha. EPLIN is a tumour suppressor protein, whose expression is altered in multiple common human tumour types. EPLIN nucleic acids and proteins are used in screening assays to detect molecules that specifically bind to EPLIN nucleic acids, proteins or derivatives and thus have potential use as agonist or antagonist of EPLIN, in particular molecules that affect cell proliferation. Thus the assays are useful for screening molecules with potential utility as anticancer drugs or lead compounds for drug development. EPLIN nucleic acids, proteins are useful for detecting a cell proliferative disorder in a subject. EPLIN polynucleotides are useful in gene therapy techniques. EPLIN is useful as a marker that can be diagnostically, prognostically and therapeutically used over the course of a cell proliferative disorder associated with EPLIN.

Sequence 759 AA;
 Query Match 89.3%; Score 678; DB 22; Length 759;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MESSPFNRRQWTSLSRVTAKEISLVNKNKSSAIVEIFSKYQAAETNMKRSNTENL 60
 Db 1 MESSPFNRRQWTSLSRVTAKEISLVNKNKSSAIVEIFSKYQAAETNMKRSNTENL 60
 Qy 61 SQHFRKGTTLVLKKKNWPNGLGAESHTDLSRNSSTEIRHRADHPAETVSHAASGAKAQ 120
 Db 61 SQHFRKGTTLVLKKKNWPNGLGAESHTDLSRNSSTEIRHRADHPAETVSHAASGAKAQ 120
 Qy 121 BEQIHPSRLRSPEALVQGRPHIKDGEDLKHSTESKVENCLGESRHEVEKSEISEN 180
 Db 121 BEQIHPSRLRSPEALVQGRPHIKDGEDLKHSTESKVENCLGESRHEVEKSEISEN 180
 Qy 181 TDASGKIEKNVPLNRLKMWFEKGEPTQTKILRAQSRASGRKISENSYSLDDLEIGPQ 240
 Db 181 TDASGKIEKNVPLNRLKMWFEKGEPTQTKILRAQSRASGRKISENSYSLDDLEIGPQ 240

Qy 241 LSSSTFDSEKVESRRNLELPRLSETSIKORMAKYQAAVSKQSSSTNTYNELKASGGEIKI 300
 Db 241 LSSSTFDSEKVESRRNLELPRLSETSIKORMAKYQAAVSKQSSSTNTYNELKASGGEIKI 300
 Qy 301 HKMEQKENVPPGPEVCITHOEGEKISANENSLAVRSTPAEDDSDRDSDQVSKSEVQPVHPKP 360
 Db 301 HKMEQKENVPPGPEVCITHOEGEKISANENSLAVRSTPAEDDSDRDSDQVSKSEVQPVHPKP 360
 Qy 361 LSPDSRASSLSSESSPPKAMKKFOAPARETCVBCQKTVYPMERLLANQQVFHISCFCSCSYC 420
 Db 361 LSPDSRASSLSSESSPPKAMKKFOAPARETCVBCQKTVYPMERLLANQQVFHISCFCSCSYC 420
 Qy 421 NNKLSLGTYSALHGRYICKPHFNQPKSGNVDGFGHPRHDKDLWASKNENESILERPAP 480
 Db 421 NNKLSLGTYSALHGRYICKPHFNQPKSGNVDGFGHPRHDKDLWASKNENESILERPAP 480
 Qy 481 LANARETHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRIAPPPPTLGS 540
 Db 481 LANARETHSPGVEDAPIAKVGLAASMEAKASSQOEKEDKPAETKKLRIAPPPPTLGS 540
 Qy 541 SGSALBEGIKMKSPKWPPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPFTVAASFQST 600
 Db 541 SGSALBEGIKMKSPKWPPEDEISKPEVPEDVDLDLKLRRSSSLKERSRPFTVAASFQST 600
 Qy 601 SVKSPKTVSPPIRKGMSEGESEESVGRVARKQVENAKASKNGNKGKTTWQNKESKG 660
 Db 601 SVKSPKTVSPPIRKGMSEGESEESVGRVARKQVENAKASKNGNKGKTTWQNKESKG 660
 Qy 661 ETGKRSKEGHSLEMEENEN 678
 Db 661 ETGKRSKEGHSLEMEENEN 678

RESULT 5
 AAB42934
 ID AAB42934 standard; Protein; 760 AA.
 XX
 AC AAB42934;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2698 polypeptide sequence SEQ ID NO:5396.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipruritic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200038473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.

Shinkets RA, Leach M;
WPI; 2000-602362/57.
N-PSDB; AAC7143.

Novel nucleic acids and peptides derived from open reading frame X,
useful for treating e.g. cancers, proliferative disorders,
neurodegenerative disorders and cardiovascular disease -

Claim 11; Page 4580-4582; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
which represent the human ORFX open reading frames 1 to 3161. The ORFX
sequences have activities such as: cytostatic; hepatotropic; vulnery;
antiporatic; antiparkinsonian; niotropic; neuroprotective;
osceoparatic; anticonvulsant; antiarthritic; immunosuppressant;
immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
antidiabetic; hypotensive; dermatological; immunosuppressive;
antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
antithyroid; and antianemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an ORFX-associated disorder. The
nucleic acids can be used to express ORFX proteins in gene therapy
vectors. The proteins and nucleic acids may be used to treat cancers,
proliferative disorders, neurodegenerative disorders, osteoarthritis,
graft vs host disease, cardiovascular disease, diabetes mellitus,
hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
bacterial or fungal infection, malaria, autoimmune disorders, asthma,
allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
nocturnal haemoglobinuria, antiinflammatory disease; to enhance
coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 760 AA;

Query Match 70.9%; Score 538; DB 21; Length 760;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 758; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MESSPFNRRQWTSLSLRTAKEISLVNKNSSAIVEFSYQKAAEETNMKKSNTEYL 60
Db 1 MESSPFNRRQWTSLSLRTAKEISLVNKNSSAIVEFSYQKAAEETNMKKSNTEYL 60

QY 61 SQHFRKGTLLVKKWENPGLGAESHTDSLNSSTEIRHADHPPEAVTSHAASCAKADQ 120
Db 61 SQHFRKGTLLVKKWENPGLGAESHTDSLNSSTEIRHADHPPEAVTSHAASCAKADQ 120

QY 121 EQGIHPRSRLSPPEALVQGRYPHIKQGEDLKHSTESKKWENCLGSRHEVEKSEISEN 180
Db 121 EQGIHPRSRLSPPEALVQGRYPHIKQGEDLKHSTESKKWENCLGSRHEVEKSEISEN 180

QY 181 TDASGKTEKNVPLNRLKWMFEGEPTQTILRAQSRASGRKISENSYSLDLEIGPGQ 240
Db 181 TDASGKTEKNVPLNRLKWMFEGEPTQTILRAQSRASGRKISENSYSLDLEIGPGQ 240

QY 241 LSSSTFDSEKNESRRNLELPRLETSIKDRMAKYQAAVSKQSSSTNTYNELKASGGI 300
Db 241 LSSSTFDSEKNESRRNLELPRLETSIKDRMAKYQAAVSKQSSSTNTYNELKASGGI 300

QY 301 HKMEQKENVPPGPEVCITHOEGEKISANENS LAVRSTPAEDDS -RDSQVKSEVQOPVHPK 359
Db 301 HKMEQKENVPPGPEVCITHOEGEKISANENS LAVRSTPAEDDSPGDSQVKSEVQOPVHPK 360

QY 360 PLSPDGRASSLSSESSPPKAWKKQAPARETCVBCQKTVYPMERLLANQQVFHISCFRC 419
Db 361 PLSPDGRASSLSSESSPPKAWKKQAPARETCVBCQKTVYPMERLLANQQVFHISCFRC 420

QY 420 CNNKLSIGTVASLHGRITYCKPHFNQLFKSKGNVDEGFGHRPHDXDLWASKNENEIILRPA 479
Db 421 CNNKLSIGTVASLHGRITYCKPHFNQLFKSKGNVDEGFGHRPHDXDLWASKNENEIILRPA 480

QY 480 QLANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKXLRIAWPPPT 539

Db	481	QLANARETPHSPGVEDAPIAKVGVLAAWEAKASSQKEKDPATETKKLRIAWPPTEL	G	540
Qy	540	SSGSALBEGIKWSPKPPPEDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFOS	S	599
Db	541	SSGSALBEGIKWSPKPPPEDEISKPEVPEDVDLDLKKLRSSSLKERSRPTVAASFOS	S	600
Qy	600	TSVSKPKTVPPIEKGMSSQSESVGGRVAERKQVENAKSKNGNVGKTTWQNKESK	S	659
Db	601	TSVSKPKTVPPIIRKGMSSQSESVGGRVAERKQVENAKSKNGNVGKTTWQNKESK	S	660
Qy	660	GETGKRKSGHLSLEVENLVENGADSDDEDNSFLKQOSPQEPKSLNWSFVDNTFAEEF	S	719
Db	661	GETGKRKSGHLSLEVENLVENGADSDDEDNSFLKQOSPQEPKSLNWSFVDNTFAEEF	S	720
Qy	720	TTQNKQSDVELWEGEVVYKELSVBEQIKRNRYYDEDEEE	S	759
Db	721	TTQNKQSDVELWEGEVVYKELSVBEQIKRNRYYDEDEEE	S	760
RESULT	6			
AA	95477			
ID	AA895477	standard; Protein; 457 AA.		
XX	AA895477;			
XX	26-JUN-2001	(first entry)		
XX	Human protein sequence	SEQ ID NO:17592.		
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy.			
XX	Homo sapiens.			
XX	EP1074617-A2.			
XX	07-FEB-2001.			
XX	28-JUL-2000; 2000EP-0116126.			
XX	29-JUL-1999; 99JP-0248036.			
XX	27-AUG-1999; 99JP-0300253.			
XX	11-JAN-2000; 2000JP-0118776.			
XX	02-MAY-2000; 2000JP-0183767.			
XX	09-JUN-2000; 2000JP-0241899.			
XX	(HELI-) HELIX RES INST.			
XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;			
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;			
XX	WPI; 2001-318749/34.			
XX	Primer sets for synthesizing polynucleotides, particularly the 5602			
XX	full-length cDNAs defined in the specification, and for the detection			
XX	and/or diagnosis of the abnormality of the proteins encoded by the			
XX	full-length cDNAs -			
XX	Claim 8; SEQ ID 17992; 2537pp + CD ROM; English.			
XX	The present invention describes primer sets for synthesizing 5602			
XX	full-length cDNAs defined in the specification. Where a primer set			
XX	comprises: (a) an oligo-dr primer and an oligonucleotide complementary			
XX	to the complementary strand of a polynucleotide which comprises one of			
XX	the 5602 nucleotide sequences defined in the specification, where the			
XX	oligonucleotide comprises at least 15 nucleotides; or (b) a combination			
XX	of an oligonucleotide comprising a sequence complementary to the			
XX	complementary strand of a polynucleotide which comprises a 5'-end			
XX	sequence and an oligonucleotide comprising a sequence complementary to a			
XX	polynucleotide which comprises a 3'-end sequence, where the			
XX	oligonucleotide comprises at least 15 nucleotides and the combination of			
XX	the 5'-end sequence/3'-end sequence is selected from those defined in			
XX	the specification. The primer sets can be used in antisense therapy and			

CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX SQ Sequence 457 AA;

Query Match 60.2%; Score 457; DB 22; Length 457;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 MEQKENVPPGVCVTHOGEKISANENSLAVRSTPAEDDSDRSQVVKPKPLS 362
 Db 1 MEQKENVPPGVCVTHOGEKISANENSLAVRSTPAEDDSDRSQVVKPKPLS 60
 QY 363 PDSRASSLSSESPPKAMKKFOAPARETCVCEQKTVYPMERLLANQOVPHISCFRCSCYNN 422
 Db 61 PDSRASSLSSESPPKAMKKFOAPARETCVCEQKTVYPMERLLANQOVPHISCFRCSCYNN 120
 QY 423 KLSLGTYSALHGRVYCKPHFNQLFKSKGNVDEGFGHRPHKOLWASKNENEILRRPAQLA 482
 Db 121 KLSLGTYSALHGRVYCKPHFNQLFKSKGNVDEGFGHRPHKOLWASKNENEILRRPAQLA 180
 QY 483 NARETPHSGVEDAPIAKVGLAASMEAKASSQEKEDKPAETKKLRIAMPPTTELSSG 542
 Db 181 NARETPHSGVEDAPIAKVGLAASMEAKASSQEKEDKPAETKKLRIAMPPTTELSSG 240
 QY 543 SALEEGIKMSKPKPPPEDEISKPEVPEDVDLDLKLRRSSSLKRSRPTVAASFQSTSV 602
 Db 241 SALEEGIKMSKPKPPPEDEISKPEVPEDVDLDLKLRRSSSLKRSRPTVAASFQSTSV 300
 QY 603 KSPKTVSPPIRKGNMSQSESVGGRVAERKQOVENAKASKKNGNVGKTTWQNKESKGET 662
 Db 301 KSPKTVSPPIRKGNMSQSESVGGRVAERKQOVENAKASKKNGNVGKTTWQNKESKGET 360
 QY 663 GKRSKEGHSLEMEENLVENGADSDSDNSFLKQSQPQPKSLNWSFVONTFAEFTTQ 722
 Db 361 GKRSKEGHSLEMEENLVENGADSDSDNSFLKQSQPQPKSLNWSFVONTFAEFTTQ 420
 QY 723 NQKSDVELMEGEVVKLSVEEQIKRNYRYDEDEE 759
 Db 421 NQKSDVELMEGEVVKLSVEEQIKRNYRYDEDEE 457

RESULT 7

AAH48487
 ID AAY48487 standard; Protein; 471 AA.

XX AC AAY48487;

XX DT 08-DEC-1999 (first entry)

XX DE Human breast tumour-associated protein 32.

XX KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;
 XX KW medicaments; gene therapy; treatment; fat metabolism.

OS Homo sapiens.

XX PN DE19813835-A1

XX PD 23-SEP-1999.

XX PF 20-MAR-1998; 98DE-1013835.

XX PR 20-MAR-1998; 98DE-1013835.

XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
 XX WPI; 1999-528979/45.
 DR N-PSDB; AAZ33566.
 XX Human nucleic acid sequences and protein products from normal breast
 PT tissue, useful for breast cancer therapy
 XX Claim 28; 173; 206pp; German.
 XX This invention describes novel human nucleic acid sequences from normal
 CC breast tissue which have cytostatic activity. The nucleic acid sequences
 CC can be used to produce and isolate full-length gene sequences. They can
 CC be used to express proteins, which can be used as tools to find an
 CC activity against breast cancer. The sequences can be used in sense or
 CC antisense form. They are especially useful for medicaments for gene
 CC therapy to treat breast cancer and for treating illnesses associated
 CC with fat metabolism. AAY48456-48539 represent protein fragments encoded
 CC by the expressed sequence tags described in the method of the invention.
 XX SQ Sequence 471 AA;

Query Match 48.7%; Score 370; DB 20; Length 471;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 470; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 289 NELKASGGEIKHMKQENVPKPPKAMKKFOAPARETCVCEQKTVYPMERLLANQO 348
 Db 1 NELKASGGEIKHMKQENVPKPPKAMKKFOAPARETCVCEQKTVYPMERLLANQO 60
 QY 349 KSEVQOPVHPKPLSPDSRASSLSSESPPKAMKKFOAPARETCVCEQKTVYPMERLLANQO 408
 Db 61 KSEVQOPVHPKPLSPDSRASSLSSESPPKAMKKFOAPARETCVCEQKTVYPMERLLANQO 120
 QY 409 VFHISCFRCSCYNNKLSLGTYSALHGRVYCKPHFNQLFKSKGNVDEGFGHRPHKOLWASK 468
 Db 121 VFHISCFRCSCYNNKLSLGTYSALHGRVYCKPHFNQLFKSKGNVDEGFGHRPHKOLWASK 180
 QY 469 NENEEILRRPAQLANARETPHSGVEDAPIAKVGLAASMEAKASSQEKEDKPAETKKL 528
 Db 181 NENEEILRRPAQLANARETPHSGVEDAPIAKVGLAASMEAKASSQEKEDKPAETKKL 240
 QY 529 RIAMPPTTELSSGSALEEGIKMSKPKPPPEDEISKPEVPEDVDLDLKLRRSSSLKERS 588
 Db 241 RIAMPPTTELSSGSALEEGIKMSKPKPPPEDEISKPEVPEDVDLDLKLRRSSSLKERS 300
 QY 589 RPFTVAASFQSTSVKSPKTVSPPIRKGNMSQSESVGGRVAERKQOVENAKASKKNGNV 648
 Db 301 RPFTVAASFQSTSVKSPKTVSPPIRKGNMSQSESVGGRVAERKQOVENAKASKKNGNV 360
 QY 649 GKTITWQNKESKGETGKRSKEGHSLEMEENLVENGADSDSDNSFLKQSQPQPKSLNWS 708
 Db 361 GKTITWQNKESKGETGKRSKEGHSLEMEENLVENGADSDSDNSFLKQSQPQPKSLNWS 420
 QY 709 SFVDNTFAEFTTQNKSDVELMEGEVVKLSVEEQIKRNYRYDEDEE 759
 Db 421 SFVDNTFAEFTTQNKSDVELMEGEVVKLSVEEQIKRNYRYDEDEE 471

RESULT 8

AAB67700

ID AAB67700 standard; Protein; 600 AA.

XX AC AAB67700;

XX DT 11-JUN-2001 (first entry)

XX DE Human EPLIN (epithelial protein lost in neoplasm)-alpha isoform.
 XX KW Human; EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;
 KW EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;
 KW gene therapy; cancer.

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 118 /note= "Val encoded by GCG"
XX FT Misc-difference 336 /note= "Asp encoded by AAT"
XX FT Misc-difference 352 /note= "Lys encoded by AGG"
XX FT Misc-difference 463 /note= "Ser encoded by AAT"
XX FT Misc-difference 499 /note= "Ser encoded by TTT"
XX FT Misc-difference 520 /note= "Ile encoded by TTT"
XX FT Misc-difference 539 /note= "Ser encoded by TTT"
XX FT Misc-difference 545 /note= "Ser encoded by TTT"
XX PN WO200118019-A1.
XX PD 15-MAR-2001.
XX PF 08-SEP-2000; 2000WO-US24689.
XX PR 08-SEP-1999; 99US-0153024.
XX PA (REGC) UNIV CALIFORNIA.
XX PI Chang DD, Maul RS;
XX PI WPI; 2001-244555/25.
XX DR N-PSDB; AAF5696.
XX PT New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
XX PT prognostic and therapeutic applications over the course of cell
XX PT proliferative disorders associated with EPLIN -
XX PS Claim 1; Page 43; 59pp; English.
XX CC The present sequence represents a human EPLIN (epithelial protein lost in
XX CC neoplasm)-alpha isoform. The specification also describes EPLIN-beta.
XX CC EPLIN is a tumour suppressor protein, whose expression is altered in
XX CC multiple common human tumour types. EPLIN nucleic acids and proteins are
XX CC used in screening assays to detect molecules that specifically bind to
XX CC EPLIN nucleic acids, proteins or derivatives and thus have potential use
XX CC as agonist or antagonist of EPLIN, in particular molecules that affect
XX CC cell proliferation. Thus the assays are useful for screening molecules
XX CC with potential utility as anticancer drugs or lead compounds for drug
XX CC development. EPLIN nucleic acids, proteins are useful for detecting a
XX CC cell proliferative disorder in a subject. EPLIN polynucleotides are
XX CC useful in gene therapy techniques. EPLIN is useful as a marker that
XX CC can be diagnostically, prognostically and therapeutically used over
XX CC the course of a cell proliferative disorder associated with EPLIN.
XX SQ Sequence 600 AA;
XX Query Match 44.0%; Score 334; DB 22; Length 600;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 334; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 345 DSQKSEVQVPHKPLSPDSRASSLSESSPPKAMKXFQAPARETCVCOKTVYPMERLL 404
Db 186 DSQKSEVQVPHKPLSPDSRASSLSESSPPKAMKXFQAPARETCVCOKTVYPMERLL 245
Oy 405 ANQOVFHSFCRCYCYNKLSLGTYSLSHGRYICKPHFNQLFKSGNYDSGFGHRPHKOL 464
Db 246 ANQOVFHSFCRCYCYNKLSLGTYSLSHGRYICKPHFNQLFKSGNYDSGFGHRPHKOL 305
Oy 465 WASKNENEILERPAQLANARETTPHSPGVEDAPIAKVGVLAASWEAKASSQOQEKDPAE 524

Db 306 WASKNENEILERPAQLANARETTPHSPGVEDAPIAKVGVLAASWEAKASSQOQEKDPAE 365
Oy 525 TKKLRIAWPPPTTELGGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLKLRRSSSL 584
Db 366 TKKLRIAWPPPTTELGGSSGSALEEGIKMSKPKWPPPEDEISKPEVPEDVDLKLRRSSSL 425
Oy 585 KERSRPFTVAASFOSTSVKSPKTVSPPIRGKMSVSESESVGGRVAERKOVENAKASKK 644
Db 426 KERSRPFTVAASFOSTSVKSPKTVSPPIRGKMSVSESESVGGRVAERKOVENAKASKK 485
Oy 645 NGNVGKTTWQNKESKGETGKRSKEGHSLENNEN 678
Db 486 NGNVGKTTWQNKESKGETGKRSKEGHSLENNEN 519
RESULT 9
AAB34206
ID AAB34206 standard; Protein; 299 AA.
XX AC AAB34206;
XX DT 26-JAN-2001 (first entry)
XX DE Human secreted protein sequence encoded by gene 41 SEQ ID NO:174.
XX KW Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
XX KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
XX KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
XX KW fungicide; ophthalmological; vulnerary; gene therapy; autoimmune disease;
XX KW hyperproliferative disorder; neoplasm; cancer; cardiovascular disorder;
XX KW cerebrovascular disorder; angiogenesis; nervous system disorder;
XX KW infection; ocular disorder; wound healing; skin aging; food additive;
XX KW preservative.
XX OS Homo sapiens.
XX PN WO200056755-A1.
XX PD 28-SEP-2000.
XX PF 16-MAR-2000; 2000WO-US06830.
XX PR 19-MAR-1999; 99US-0125361.
XX PR 10-DEC-1999; 99US-0169910.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SW, Komatsoulis G;
XX PI WPI; 2000-587661/55.
XX DR New isolated nucleic acid molecules encoding 49 human secreted proteins
XX FT used for preventing, treating or ameliorating medical conditions, for
XX FT diagnosing pathological conditions or as food additives or
XX FT preservatives -
XX PS Disclosure; Page 412-413; 419pp; English.
XX CC The polynucleotide sequences given in AAC59449 to AAC59497 encode the
XX CC human secreted proteins given in AAB34092 to AAB34140. AAB34141 to
XX CC AAB34216 represent human secreted polypeptide sequences and proteins
XX CC homologous to them, which are given in the exemplification of the present
XX CC invention. Human secreted proteins have activities based on the tissue
XX CC and cells the genes are expressed in. Examples of activities include:
XX CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
XX CC cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;
XX CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX CC and vulnerary. The polynucleotides and polypeptides can be used to
XX CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX CC in diagnosing a pathological condition or susceptibility to a
XX CC pathological condition. Disorders which are diagnosed or treated include
XX CC autoimmune diseases, hyperproliferative disorders e.g. neoplasms or

CC cancer of the breast or liver, cardiovascular disorders, cerebrovascular
CC disorders, angiogenesis, nervous system disorders, infections caused by
CC bacteria, viruses and fungi and ocular disorders. The polypeptides can
CC also be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used
CC in the exemplification of the present invention.

XX Sequence 299 AA;

Query Match 39.4%; Score 299; DB 21; Length 299;
Best Local Similarity 100.0%; Pred. No. 6.6e-286; Length 299;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 EGFGRPHKDLWASKNNEEILERPALANARETPHSPGVEDAPIAKVGVLAASMEAKAS 513

Db 1 EGFGRPHKDLWASKNNEEILERPALANARETPHSPGVEDAPIAKVGVLAASMEAKAS 60

QY 514 SQQEKEDKPAETKKLRITAWPPPTLGGSSGSALEGIKMSKPKWPPDEISKPEVPEDVDL 573

Db 61 SQQEKEDKPAETKKLRITAWPPPTLGGSSGSALEGIKMSKPKWPPDEISKPEVPEDVDL 120

QY 574 DLKKLRSSSLKERSRPTVAASFOSTSVKSPKTVSPPIKKGWSMSQSESVGGRVAER 633

Db 121 DLKKLRSSSLKERSRPTVAASFOSTSVKSPKTVSPPIKKGWSMSQSESVGGRVAER 180

QY 634 KOVENAKASKKNGVNGKTTWQNKESKGETGKRSGEHSLEMEENLVENGADSDDDNSF 693

Db 181 KOVENAKASKKNGVNGKTTWQNKESKGETGKRSGEHSLEMEENLVENGADSDDDNSF 240

QY 694 LKQSPQEPKSLNWSFVNTFAEFTTQNKQGDVELWGEVVKLSVEEQIKRNRYY 752

Db 241 LKQSPQEPKSLNWSFVNTFAEFTTQNKQGDVELWGEVVKLSVEEQIKRNRYY 299

RESULT 10

ABBS6420

ID ABB56420 standard; Protein; 301 AA.

XX AC ABB56420;

DT 21-FEB-2002 (first entry)

XX Human cancer suppressor protein PP624.

XX Human; cancer suppressor; disease; cancer.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 21

FT /label= unknown

FT /note= "Encoded by RCA"

FT Misc-difference 40

FT /label= unknown

FT /note= "Encoded by TRC"

FT Misc-difference 43

FT /label= unknown

FT /note= "Encoded by AKC"

XX CN1313297-A.

PN 19-SEP-2001.

XX 09-MAR-2000; 2000CN-0111948.

XX 09-MAR-2000; 2000CN-0111948.

XX (SHAN-) SHANGHAI INST ONCOLOGY.

XX Gu J, Yang S;
PI WPI; 2002-042185/06.
XX N-PSDB; ABI98973.
XX Human protein able to suppress growth of cancer cells and its coding
PT sequence -
XX Claim 1; Page 20 Disclosure; 37pp; Chinese.
XX The invention relates to novel human proteins (ABB56417-ABB56425) with
CC cancer suppressing function, the encoding polynucleotides
CC (ABI98970-ABI98978), the process for preparing the polypeptide, the
CC application of the polypeptide in treating diseases such as cancer, the
CC antagonist of the polypeptide and its medical function and the
CC application of the polynucleotide.
SQ Sequence 301 AA;

Query Match 31.6%; Score 240; DB 23; Length 301;
Best Local Similarity 100.0%; Pred. No. 1e-227;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 NKLSLGTVASLHGRYCKPHFNQLFKSGNYDEGFGRPHKDLWASKNNEEILERPQOL 481

Db 44 NKLSLGTVASLHGRYCKPHFNQLFKSGNYDEGFGRPHKDLWASKNNEEILERPQOL 103

QY 482 ANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAMPPPTLGLSS 541

Db 104 ANARETPHSPGVEDAPIAKVGVLAASMEAKASSQOEKEDKPAETKKLRIAMPPPTLGLSS 163

QY 542 GSALGEGIKMSKPKWPPDEISKPEVPEDVDLKLRRSSSLKERSRPTVAASFOSTS 601

Db 164 GSALGEGIKMSKPKWPPDEISKPEVPEDVDLKLRRSSSLKERSRPTVAASFOSTS 223

QY 602 VKSPKTVSPPIKKGWSMSQSESVGGRVAERKOVENAKASKKNGVNGKTTWQNKESKGE 661

Db 224 VKSPKTVSPPIKKGWSMSQSESVGGRVAERKOVENAKASKKNGVNGKTTWQNKESKGE 283

RESULT 11

AAU10979

ID AAU10979 standard; Protein; 457 AA.

XX AC AAU10979;

DT 12-MAR-2002 (first entry)

XX Human sterol regulatory element binding protein 3.

XX Human; sterol regulatory element binding protein 3; hSRBP-3;
KW hypothalamus.

XX Homo sapiens.

XX CN1309182-A.

XX 22-AUG-2001.

XX 17-FEB-2000; 2000CN-0111698.

XX 17-FEB-2000; 2000CN-0111698.

XX (NATU-) NATURAL HUMAN GENOME NANFANG RES CENT.

XX Li Y, Xu S, Ren S;

XX WPI; 2002-011822/02.

XX N-PSDB; AAS18588.

XX Cholesterol regulatory factor binding protein and its coding sequence -

PS Claim 2; Page 14; 27pp; Chinese.

XX The invention relates to a novel human sterol regulatory element binding

CC protein 3 (hsREBP-3) expressed in human normal hypothalamic tissue and

CC its coding sequence. Also described is the process for preparing the

CC protein and nucleic acid sequence, and the method for detecting hsREBP-3

CC nucleic acid sequence and polypeptides. The present sequence represents

CC the amino acid sequence of human hsREBP-3 as described in the invention.

XX

SQ Sequence 457 AA;

Query Match 31.4%; Score 238; DB 23; Length 457;

Best Local Similarity 100.0%; Pred. No. 1.4e-225; Indels 0; Gaps 0;

Matches 238; Conservative 0; Mismatches 0;

QY 522 PATETKRLIAMPPTTELSSGSALEEGIKMSKPKWPPDEISKPEVPDVLDKLRRS 581

Db 220 PATETKRLIAMPPTTELSSGSALEEGIKMSKPKWPPDEISKPEVPDVLDKLRRS 279

QY 582 SSLKERSPFTVAASFQSTSVKSPKTVPPPIRKGWSMSQSEESVGGVRAERKQVENAKA 641

Db 280 SSLKERSPFTVAASFQSTSVKSPKTVPPPIRKGWSMSQSEESVGGVRAERKQVENAKA 339

QY 642 SKKNGNVGKTTWQNKESKGETGKRSKEGHSLEMEENLVENGADSDDDNSFLKQSQPQE 701

Db 340 SKKNGNVGKTTWQNKESKGETGKRSKEGHSLEMEENLVENGADSDDDNSFLKQSQPQE 399

QY 702 PKSLNWSFVNTFAEFTTQNKSQDVVELWEGVVKLSVEEQIKRNYRYDEDEEE 759

Db 400 PKSLNWSFVNTFAEFTTQNKSQDVVELWEGVVKLSVEEQIKRNYRYDEDEEE 457

RESULT 12

AAB51883

ID AAB51883 standard; Protein; 232 AA.

XX

AC AAB51883;

XX

DT 16-FEB-2001 (first entry)

XX

DE Human secreted protein sequence encoded by gene 5 SEQ ID NO:116.

XX

KW Human; secreted protein; immunosuppressive; antiarthritic; anirheumatic;

KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;

KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;

KW ophthalmological; vulnery; autoimmune disease; rheumatoid arthritis;

KW hyperproliferative disorders; cancer; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; nervous system disorder;

KW Alzheimer's disease; ocular disorder; wound healing; skin aging.

XX

OS Homo sapiens.

XX

PN WO200061626-A1.

XX

PD 19-OCT-2000.

XX

EF 06-APR-2000; 2000WO-US09066.

XX

PR 09-APR-1999; 99US-0128698.

PR 20-JAN-2000; 2000US-0176926.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

XX

XX Rosen CA, Ruben SM, Komatsoulis G;

XX

XX WPI; 2000-619227/59.

DR N-PSDB; AAC9483.

XX

XX New nucleic acid molecules encoding 49 human secreted proteins for

PT diagnosing, preventing or ameliorating medical conditions and used for

PT food additives or preservatives -

XX

PS Disclosure; Page 22; 516pp; English.

XX Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding

CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -

CC AAB51927 represent alternative polypeptides encoded by the genes, and

CC amino acid sequences with which they share homology. The genes and

CC proteins have activities dependent on the tissues and cells in which they

CC are expressed. Examples of their activities include immunosuppressive;

CC antiarthritic; anirheumatic; antiproliferative; cytostatic; cardiant;

CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;

CC virucide; fungicide; ophthalmological; and vulnerary. The secreted

CC proteins, polynucleotides, antagonists and agonists may be useful in

CC treating, preventing and/or diagnosing diseases and disorders such as

CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative

CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders

CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,

CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,

CC infections caused by bacteria, viruses and fungi and ocular disorders

CC e.g. corneal infection. The polypeptides can also be used to aid wound

CC healing and epithelial cell proliferation, to prevent skin aging due to

CC sunburn, to maintain organs before transplantation, for supporting cell

CC culture of primary tissues, to regenerate tissues and in chemotaxis. The

CC polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities, fat content, lipid, protein,

CC carbohydrate, vitamins, minerals, cofactors and other nutritional

CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are

CC used in the isolation and characterisation of the proteins and

CC polynucleotides of the invention.

XX

SQ Sequence 232 AA;

Query Match 29.9%; Score 227; DB 21; Length 232;

Best Local Similarity 100.0%; Pred. No. 5.4e-215; Indels 0; Gaps 0;

Matches 227; Conservative 0; Mismatches 0;

QY 473 EILERPAQLANARETPHSPGVEDAPIAKGVGLAAWEAKASSQOEKEDKPAETKRLIAW 532

Db 1 EILERPAQLANARETPHSPGVEDAPIAKGVGLAAWEAKASSQOEKEDKPAETKRLIAW 60

QY 533 PPPTLGGSSGSALEEGIKMSKPKWPPDEISKPEVPDVLDKLRRSSSLKRSRPPT 592

Db 61 PPPTLGGSSGSALEEGIKMSKPKWPPDEISKPEVPDVLDKLRRSSSLKRSRPPT 120

QY 593 VAASFQSTSVKSPKTVSPPIRKGWSMSQSEESVGGVRAERKQVENAKSKNGNVGKTT 652

Db 121 VAASFQSTSVKSPKTVSPPIRKGWSMSQSEESVGGVRAERKQVENAKSKNGNVGKTT 180

QY 653 WQNKESKGETGKRSKEGHSLEMEENLVENGADSDDDNSFLKQSQP 699

Db 181 WQNKESKGETGKRSKEGHSLEMEENLVENGADSDDDNSFLKQSQP 227

RESULT 13

AAB54159

ID AAB54159 standard; Protein; 243 AA.

XX

AC AAB54159;

XX

DT 09-MAR-2001 (first entry)

XX

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:611.

XX

KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;

KW detection; diagnosis; identification; cytostatic; neuroprotective;

KW neurotropic; immunomodulatory; relaxant; contraceptive; gynaecological;

KW antiinflammatory; cardiant; gene therapy; chromosome mapping;

KW linkage analysis; tissue identification; tissue typing; forensic;

KW neural; immune system; muscular; reproductive; gastrointestinal;

KW pulmonary; cardiovascular; renal; proliferative.

XX

OS Homo sapiens.

XX

PN WO200055320-A1.

XX 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US05989.
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI: 2000-579444/54.
XX N-PSDB; AAC98924.
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
XX treating, or ameliorating a medical condition, particular pancreatic
XX cancer, or for use in assays for diagnosing a pathological condition -
XX Claim 11; Page 1050-1051; 1379pp; English.
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
XX proteins, called pancreatic cancer antigens, given in AAB54008 to
XX AAB54466. The human pancreatic cancer antigens have cytostatic,
XX neuroprotective, nontoxic, immunomodulatory, relaxant, contraceptive,
XX gynaecological, cardiac and antiinflammatory activities, and can be used
XX in gene therapy. The polynucleotide and proteins can be used for
XX preventing, treating, or ameliorating a medical condition or in assays
XX for diagnosing a pathological condition or a susceptibility to one in a
XX subject. Binding partners to the proteins and the activity of the
XX proteins can be identified. The pancreatic cancer antigens can be used to
XX detect, treat or prevent pancreatic disorders, especially cancer.
XX Agonists and antagonists to the antigens can be screened for. The
XX pancreatic cancer antigen polynucleotides can be used to design nucleic
XX acid hybridization probes that can be used in chromosome mapping, linkage
XX analysis, tissue identification and/or typing and a variety of forensic
XX and diagnostic methods. The proteins can be used to generate antibodies
XX which are used to purify, detect and target the polypeptides, including
XX both in vivo and in vitro diagnostic and therapeutic methods. The
XX proteins can be used to treat or prevent neural, immune system, muscular,
XX reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
XX proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
XX sequences used in the exemplification of the present invention.
XX Sequence 243 AA;
Query Match 24.2%; Score 184; DB 21; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.4e-172;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 321 EGEKISANENSLAVRSTPAEDDSDSDSOVKSEVQVHPKPLSPDRASSLSSESPPKAMK 380
Db 1 EGEKISANENSLAVRSTPAEDDSDSDSOVKSEVQVHPKPLSPDRASSLSSESPPKAMK 60
QY 381 KFOAPARETCVEQKTVYPMERLLANQQVPHISCFRCSCYNNKLSLGTYSALHGRYICKP 440
Db 61 KFOAPARETCVEQKTVYPMERLLANQQVPHISCFRCSCYNNKLSLGTYSALHGRYICKP 120
QY 441 HFNLQFKSKNGYDEGCHRPKOLWASKNENEFILRPAQLANARETPHSPGVEDAPIAK 500
Db 121 HFNLQFKSKNGYDEGCHRPKOLWASKNENEFILRPAQLANARETPHSPGVEDAPIAK 180
QY 501 VGVLL 504
Db 181 VGVLL 184
RESULT 14
AAV73916
ID AAY73916 standard; Protein; 235 AA.
XX AAY73916;
XX AAY73916;
DT 14-MAR-2000 (first entry)

XX Human prostate tumor EST fragment derived protein #103.
XX Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
XX treatment.
XX Homo sapiens.
XX DE19820190-A1.
XX 04-NOV-1999.
XX 28-APR-1998; 98DE-1020190.
XX 28-APR-1998; 98DE-1020190.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX WPI: 1999-621386/54.
XX N-PSDB; AAZ52892.
XX New human nucleic acid sequences from pancreatic tumors, and related
XX proteins -
XX Claim 23; Page 351; 502pp; German.
XX This invention describes novel polypeptides and their encoding nucleic
XX acids derived from human pancreatic tumor tissue which have cytostatic
XX activity. The sequences are also useful in producing pharmaceutical
XX compositions for treatment of pancreatic tumors. AAY73814-Y74252
XX represent protein fragments encoded by the human pancreatic tumor cDNA
XX library derived expressed sequence tag (EST) sequences represented in
XX AAZ52858-Z53014.
XX Sequence 235 AA;
Query Match 22.4%; Score 170; DB 20; Length 235;
Best Local Similarity 100.0%; Pred. No. 9e-159;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MESSPFNRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYKAAEETNMEKRSNTENL 60
Db 10 MESSPFNRQWTSLSLRVTAKELSLVNKNKSSAIVEIFSKYKAAEETNMEKRSNTENL 69
QY 61 SCHFRKGLTTLVKKWENPGLGAESHTSLRNSSTEIRHRAHPPEVTSAAASGAKAQD 120
Db 70 SCHFRKGLTTLVKKWENPGLGAESHTSLRNSSTEIRHRAHPPEVTSAAASGAKAQD 129
QY 121 EQIHPRSLRSPPEALVQGRYPHIKDGEDLKDHSTESKKNENCLGESRH 170
Db 130 EQIHPRSLRSPPEALVQGRYPHIKDGEDLKDHSTESKKNENCLGESRH 179
RESULT 15
ABG24090
ID ABG24090 standard; Protein; 337 AA.
XX ABG24090;
XX AC ABG24090;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #24081.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.
PF
XX
PR 31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS88277.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 54449; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 337 AA;
SQ

Query Match 20.4%; Score 155; DB 22; Length 337;
Best Local Similarity 100.0%; Pred. No. 7.8e-144;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 NTELSQHFRKGTLLVLKXKXWENPGLGASHTDSLNSSTIHRADHPPAEVTSAAAG 115
Db 2 NTELSQHFRKGTLLVLKXKXWENPGLGASHTDSLNSSTIHRADHPPAEVTSAAAG 61

QY 116 AKADQEQIHPRSLRSPPEALVQGRYPHIKDGEDLKHSTESKXKXWENPGLGASHTDSLNSSTIHRADHPPAEVTSAAAG 175
Db 62 AKADQEQIHPRSLRSPPEALVQGRYPHIKDGEDLKHSTESKXKXWENPGLGASHTDSLNSSTIHRADHPPAEVTSAAAG 121

QY 176 EISENTDASGKIEKYNVPLNKLKMMFEKGEPTQK 210
Db 122 EISENTDASGKIEKYNVPLNKLKMMFEKGEPTQK 156

Search completed: January 6, 2004, 09:51:42
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:50:00 ; Search time 21 Seconds
(without alignments)
3475.811 Million cell updates/sec

Title: US-09-890-549-4
Perfect score: 759
Sequence: 1 MESSPNRQWTSLSLRVTA.....LSVESQIKRNRVYDEDEEE 759

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 76:.*
1: Pirl:.*
2: Pirl2:.*
3: Pirl3:.*
4: Pirl4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	1.1	122	2 E90243	hypothetical prote
2	8	1.1	231	2 T02585	hypothetical prote
3	8	1.1	234	2 G2758	hypothetical prote
4	8	1.1	388	2 D72740	hypothetical prote
5	8	1.1	403	2 S71236	probable N-acetyl
6	8	1.1	405	2 T27971	hypothetical prote
7	8	1.1	413	2 A84620	similar to hookles
8	8	1.1	415	2 S00543	site-specific DNA-
9	8	1.1	466	2 E95282	L-seryl-TRNAse se
10	8	1.1	642	1 JF0079	Fe(III) dicitrate
11	8	1.1	784	2 H83158	grase B - Aquifex
12	8	1.1	792	2 F70388	bone morphogenetic
13	8	1.1	1038	2 J38935	probable membrane
14	8	1.1	1038	2 J05527	hypothetical prote
15	8	1.1	2295	2 B71621	hypothetical prote
16	8	1.1	2484	2 T26216	dopamine- and CAMP
17	8	1.1	2607	2 A56388	periplasmic mercur
18	7	0.9	49	2 AB2271	hypothetical prote
19	7	0.9	64	2 G64010	nodulin-24 - soybe
20	7	0.9	85	2 S20505	hypothetical prote
21	7	0.9	92	2 T31048	hypothetical prote
22	7	0.9	93	2 H72542	13K transport prot
23	7	0.9	101	2 S45096	hypothetical prote
24	7	0.9	118	2 S45099	13K transport prot
25	7	0.9	118	2 T49168	hypothetical prote
26	7	0.9	121	2 C90510	hypothetical prote
27	7	0.9	124	2 T37334	hypothetical prote
28	7	0.9	134	2 T37334	Probable hexose tr
29	7	0.9	132	2 T10068	

30 7 0.9 135 2 B38253 kappa-enhancer-bin
31 7 0.9 139 2 S69458 hypothetical prote
32 7 0.9 139 2 B96500 hypothetical prote
33 7 0.9 141 2 T09581 probable high mobi
34 7 0.9 149 2 S39556 high mobility grou
35 7 0.9 153 2 A71893 hypothetical prote
36 7 0.9 159 2 C49773 ecdysone-dependent
37 7 0.9 159 2 A72104 hypothetical prote
38 7 0.9 159 2 B86518 hypothetical prote
39 7 0.9 160 2 AC2077 hypothetical prote
40 7 0.9 169 2 AH1032 hypothetical prote
41 7 0.9 177 2 A75446 hypothetical prote
42 7 0.9 178 2 AH1394 hypothetical prote
43 7 0.9 178 2 S26481 B. subtilis RNA po
44 7 0.9 181 2 F82750 calcium-binding pr
45 7 0.9 182 2 B32256 outer membrane pr
hypothetical 20K p

ALIGNMENTS

RESULT 1
E90243
hypothetical protein SSO0922 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: E90243
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-
arrett, R.A.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: E90243
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <KUR>
A:Cross-references: GB:AE006641; NID:g13814104; PIDN:AAK41204.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO0922

Query Match 1.1%; Score 8; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 IEKYNVPL 194
Db 37 IEKYNVPL 44
RESULT 2
T02585
hypothetical protein At2g39170 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein T16B24.19
C:Species: Arabidopsis thaliana [mouse-ear cross]
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 02-Feb-2001
C:Accession: T02585; A84814
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.
A:Reference number: Z14679
A:Accession: T02585
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-231 <ROU>
A:Cross-references: EMBL:AC004697; NID:g3402671; PID:g3402687
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: A84814
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-231 <STO>
A:Cross-references: GB:AE002093; NID:G3402687; PIDN:AAC28990.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g39170; T16B24.19
A:Map position: 2

Query Match
Best Local Similarity 1.1%; Score 8; DB 2; Length 231;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 626 VGGRAER 633
DB 223 VGGRAER 230

RESULT 3
G72758
hypothetical protein APE0060 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: G72758
R:Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: G72758
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <KAW>
A:Cross-references: DDBJ:AP000058; NID:G5103388; PIDN:BAA78969.1; PID:d1042745; PID:G5103388
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0060

Query Match
Best Local Similarity 1.1%; Score 8; DB 2; Length 234;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 ESVGGRVA 631
DB 199 ESVGGRVA 206

RESULT 4
D72740
hypothetical protein APE0453 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: D72740
R:Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: D72740
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-388 <KAW>
A:Cross-references: DDBJ:AP000059; NID:G5103911; PIDN:BAA79416.1; PID:G5104100
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0453
C:Superfamily: Aeropyrum pernix hypothetical protein APE0453

Query Match
Best Local Similarity 1.1%; Score 8; DB 2; Length 388;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 503 VLAASMEA 510

Db 97 VLAASMEA 104

RESULT 5
S71236
probable N-acetyltransferase (EC 2.3.1.-) hookless 1 - Arabidopsis thaliana
N:Alternate names: ethylene response protein; protein F19F18.70
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 11-Jun-1999
C:Accession: S71236; T04715
R:Jehman, A.L.; Black, R.; Ecker, J.R.
submitted to the EMBL Data Library, March 1996
A:Description: Hookless1, an ethylene response gene, is required for differential cell elongation in Arabidopsis thaliana
A:Reference number: S71236
A:Accession: S71236
A:Molecule type: mRNA
A:Residues: 1-403 <LEH>
A:Cross-references: EMBL:U50399; NID:G1277089; PID:G1277090
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15382
A:Accession: T04715
A:Molecule type: DNA
A:Residues: 1-403 <BEV>
A:Cross-references: EMBL:AL035605
A:Experimental source: cultivar Columbia; BAC clone F19F18
C:Genetics:
A:Gene: hookless1; HLS1
A:Map position: 4
A:Introns: 54/3; 113/2
A:Note: F19F18.70
C:Function:
A:Description: involved in differential cell elongation in the hypocotyl
C:Keywords: acyltransferase

Query Match
Best Local Similarity 1.1%; Score 8; DB 2; Length 403;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 NNKLSLGT 428
DB 215 NNKLSLGT 222

RESULT 6
T27971
hypothetical protein ZK675.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T27971
R:Sims, M.
submitted to the EMBL Data Library, November 1994
A:Reference number: Z20448
A:Accession: T27971
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-405 <WIL>
A:Cross-references: EMBL:Z46812; PIDN:CAA86845.1; GSPDB:GN00020; CESP:ZK675.3
A:Experimental source: clone ZK675
C:Genetics:
A:Gene: CESP:ZK675.3
A:Map position: 2
A:Introns: 43/2; 84/2; 110/3; 174/3; 233/3; 264/2; 335/3; 372/1
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK675.3

Query Match
Best Local Similarity 1.1%; Score 8; DB 2; Length 405;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 SKNENEI 474
DB 254 SKNENEI 261


```

A;Residues: 1-466 <KUR>
A;Cross-references: GB:AE006469; PIDN:AAK64663.1; PID:g14523060; GSPDB:GN00165
A;Experimental source: Strain 1021, megaplasmid pSymA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
heault, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: selsa
A;Genome: plasmid
C;Superfamily: Escherichia coli seryl-tRNA(Sec) selenium transferase selsa

Query Match 1.1% Score 8; DB 2; Length 466;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 GSGSQALE 546
|||||
Db 411 GSGSQALE 418

RESULT 10
JP0079
LIM protein kinase (EC 2.7.1.1) - chicken
C;Species: Gallus gallus (chicken)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: JP0079
R;Ohashi, K.; Tashima, J.; Tajinda, K.; Nakamura, T.; Mizuno, K.
J. Biochem. 116, 636-642, 1994
A;Title: Molecular cloning of a chicken lung cDNA encoding a novel protein kinase with N
A;Reference number: JP0079; MUID:95155277; PMID:7852284
A;Accession: JP0079
A;Molecule type: mRNA
A;Residues: 1-642 <OHA>
A;Cross-references: DDBJ:D26310; NID:g1402513; PIDN:BAA05372.1; PID:g643086
C;Comment: This protein plays an important role in the regulation of diverse cell function
C;Superfamily: LIM protein kinase; LIM metal-binding repeat homology; protein kinase hom
C;Keywords: ATP; duplication; phosphotransferase; serine/threonine/tyrosine-specific pro
F;12-63/Domain: LIM metal-binding repeat homology <LIM>
F;72-124/Domain: LIM metal-binding repeat homology <LIM2>
F;329-608/Domain: protein kinase homology <KIN>
F;337-345/Region: protein kinase ATP-binding motif
F;496-506/Region: basic

Query Match 1.1% Score 8; DB 1; Length 642;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 HISCRCFS 418
|||||
Db 34 HISCRCFS 41

RESULT 11
H83158
Fe(III) dicitrate transport protein FecA PA3901 [imported] - Pseudomonas aeruginosa (str
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83158
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bri
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83158
A;Status: preliminary
A;Molecule type: DNA

```

```

RESULT 7
A84620
Similar to hookless1 (HLS1) [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84620
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84620
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-413 <STO>
A;Cross-references: GB:AE002093; NID:g3169179; PIDN:AAC17822.1; GSPDB:GN00139
C;Genetics:
A;Gene: Atg23060
A;Map position: 2

Query Match 1.1% Score 8; DB 2; Length 413;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 NNKLSLGT 428
|||||
Db 225 NNKLSLGT 232

RESULT 8
S00543
site-specific DNA-methyltransferase (cytosine-specific) (EC 2.1.1.73) DdeI - Desulfovibrio
A;Alternate names: modification methylase DdeI
C;Species: Desulfovibrio desulfuricans
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 05-May-2000
C;Accession: S00543
R;Snyter, L.A.; Slatko, B.; Moran, L.; O'Donnell, K.H.; Brooks, J.E.
Nucleic Acids Res. 15, 8249-8256, 1987
A;Title: Nucleotide sequence of the DdeI restriction-modification system and characteriz
A;Reference number: S00542; MUID:88040453; PMID:2823226
A;Accession: S00543
A;Molecule type: DNA
A;Residues: 1-415 <SNZ>
A;Cross-references: EMBL:Y00449; NID:g40793; PIDN:CAA68505.1; PID:g40795
C;Superfamily: site-specific methyltransferase (cytosine-specific) EcoRII
C;Keywords: methyltransferase; S-adenosylmethionine

Query Match 1.1% Score 8; DB 2; Length 415;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 TVAASTQS 599
|||||
Db 296 TVAASTQS 303

RESULT 9
E95262
S-eryl-tRNASec selenium transferase (EC 2.9.1.1) [imported] - Sinorhizobium meliloti (S
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 03-Jun-2002
C;Accession: E95262
R;Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Suryski, R.; Wells, D.H.; Yeh, K.C.
roc. Natl. Acad. Sci. U.S.A. 98, 9883-9886, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A;Reference number: A95262; MUID:21396509; PMID:11481432
A;Accession: E95262
A;Status: preliminary
A;Molecule type: DNA

```

A;Residues: 1-784 <STO>
A;Cross-references: GB:AE004807; GB:AE004091; NID:g9500076; PIDN:AAG07288.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
C;Superfamily: vitamin B12 receptor; tonB-dependent receptor amino-terminal homology; to

Query Match 1.1%; Score 8; DB 2; Length 784;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 LGESRHEV 172
Db 440 LGESRHEV 447

RESULT 12
F70388
Gyrase B - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999
C;Accession: F70388
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: F70388
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-792 <AQF>
A;Cross-references: GB:AE000719; NID:g2983517; PIDN:AAC07098.1; PID:g2983522; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: gyrB
C;Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B

Query Match 1.1%; Score 8; DB 2; Length 792;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 DLDLKKLR 579
Db 486 DLDLKKLR 493

RESULT 13
I38935
bone morphogenetic protein receptor II precursor - human
N;Alternate names: activin receptor-like kinase type II; bone morphogenetic protein 4 re
N;Contains: protein kinase (EC 2.7.1.37)
C;Species: Homo sapiens (man)
C;Date: 16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change 24-Sep-1999
C;Accession: I38935; I55438; I37209
R;Kawabata, M.; Chytil, A.; Moses, H.L.
J. Biol. Chem. 270, 5625-5630, 1995
A;Title: Cloning of a novel type II serine/threonine kinase receptor through interaction
A;Reference number: A55947; MUID:95197572; PMID:7890683
A;Accession: I38935
A;Molecule type: mRNA
A;Residues: 1-1038 <KAW>
A;Cross-references: ENBL:U20165; NID:g704361; PIDN:AAC50105.1; PID:g704362
R;Nonno, T.; Ishikawa, T.; Saito, T.; Hosokawa, K.; Noji, S.; Wolsing, D.H.; Rosenbaum,
J. Biol. Chem. 270, 22522-22526, 1995
A;Title: Identification of a human type II receptor for bone morphogenetic protein-4 tha
A;Reference number: I55438; MUID:95403457; PMID:7673243
A;Accession: I55438
A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/ENE
A;Molecule type: mRNA
A;Residues: 1-1038 <NOH>
A;Cross-references: GB:D50516; NID:g907712; PIDN:BA09094.1; PID:g907713
R;Rosenzweig, B.L.; Imanura, T.; Okadome, T.; Cox, G.N.; Yamashita, H.; ten Dijke, P.; H
Proc. Natl. Acad. Sci. U.S.A. 92, 7632-7636, 1995

A;Title: Cloning and characterization of a human type II receptor for bone morphogenetic
A;Reference number: I37209; MUID:95372334; PMID:7644468
A;Accession: I37209
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-827, 'R', 829-1038 <ROS>
A;Cross-references: ENBL:Z48923; NID:g1009409; PIDN:CAA88759.1; PID:g1009410
C;Genetics:
A;Gene: GDB:BMPR2; BRK-3; T-ALK; BMPR3; BMPR-II
A;Cross-references: GDB:642243; OMIM:600799
A;Map position: 20pter-20qter
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C;Keywords: ATP; glycoprotein; phosphotransferase; receptor; transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1038/Product: bone morphogenetic protein receptor II #status predicted <MAT>
F;27-150/Domain: extracellular #status predicted <EXT>
F;151-170/Domain: transmembrane #status predicted <TRM>
F;201-508/Domain: protein kinase homology <KIN>
F;209-217/Region: protein kinase ATP-binding motif
F;55,110,126/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.1%; Score 8; DB 2; Length 1038;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 ENSLAQRS 336
Db 468 ENSLAQRS 475

RESULT 14
JC5527
bone morphogenetic protein type II receptor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 24-Sep-1999
C;Accession: JC5527
R;Beppu, H.; Minowa, O.; Miyazono, K.; Kawabata, M.
Biochem. Biophys. Res. Commun. 235, 499-504, 1997
A;Title: cDNA cloning and genomic organization of the mouse BMP type II receptor.
A;Reference number: JC5527; MUID:97350808; PMID:9207184
A;Accession: JC5527
A;Molecule type: mRNA
A;Residues: 1-1038 <BEP>
A;Cross-references: DDBJ:AF003942; NID:g2253704; PIDN:AAB63042.1; PID:g2253705
C;Comment: This protein is a serine/threonine kinase receptor that forms a heteromeric C
C;Genetics:
A;Gene: BMPR-II
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
F;1-26/Domain: signal sequence #status predicted <SIG>
F;151-174/Domain: transmembrane #status predicted <TRM>
F;201-508/Domain: protein kinase homology <KIN>
F;202-500/Domain: serine/threonine kinase #status predicted <STK>

Query Match 1.1%; Score 8; DB 2; Length 1038;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 ENSLAQRS 336
Db 468 ENSLAQRS 475

RESULT 15
B71621
probable membrane associated protein PFB0190c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71621
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Perteira, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: B71621
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2295 <GAR>
A:Cross-references: GB:AE001379; GB:AE001362; NID:g3845118; PIDN:AAC71827.1; PID:g384511
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0190c

Query Match 1.1%; Score 8; DB 2; Length 2295;
Best Local Similarity 100.0%; Pred.No. 88;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 88 DSLRNSST 95
Db 1043 DSLRNSST 1050

Search completed: January 6, 2004, 09:53:41
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:48:30 ; Search time 17 seconds
(without alignments)

2099.604 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSPFNRQWTSLSLRVTA.....LSVEEQIKRNYDEDEDEE 759

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	759	100.0	759	1	EPLI_HUMAN	Q9ubh6 homo sapien
2	69	9.1	753	1	EPLI_MOUSE	Q9etg0 mus musculus
3	8	1.1	398	1	LX53_BRARE	Q90421 brachydanio
4	8	1.1	405	1	YX63_CREEL	Q09379 caenorhabdi
5	8	1.1	415	1	MTDI_DESDN	P05302 desulfovibr
6	8	1.1	456	1	SELA_RHIME	P58226 rhizobium m
7	8	1.1	642	1	L1K2_CHICK	P53666 gallus gall
8	8	1.1	792	1	GYR2_AQUAE	Q67137 aquifex aco
9	8	1.1	1038	1	BMR2_HUMAN	Q13873 homo sapien
10	8	1.1	1038	1	BMR2_MOUSE	Q35607 mus musculus
11	7	0.9	85	1	YGFY_HAEIN	P44025 haemophilus
12	7	0.9	101	1	SUI1_AERPE	Q9vbg9 aeropyrum p
13	7	0.9	149	1	HMGL_VICFA	P40620 vicia faba
14	7	0.9	156	1	PYR1_METAC	Q8thl3 methanosarc
15	7	0.9	156	1	PYR1_METMA	Q8pxk6 methanosarc
16	7	0.9	159	1	CUP9_DROME	P27781 drosophila
17	7	0.9	178	1	RPOE_LISMO	Q8y494 listeria mo
18	7	0.9	195	1	GRPE_FRATU	P48204 francisella
19	7	0.9	198	1	GRPE_BACSH	Q69267 bacillus sp
20	7	0.9	198	1	VH23_SFVKA	Q9q907 shope fibro
21	7	0.9	203	1	YPT2_WAIZE	Q25737 zea mays (m
22	7	0.9	214	1	IPYR_ORISA	Q22537 oryza sativ
23	7	0.9	215	1	IPYR_HORVD	Q23979 hordeum vul
24	7	0.9	228	1	THIE_METAC	Q8lmd6 methanosarc
25	7	0.9	233	1	CYCX_NITEU	Q50926 nitrosomona
26	7	0.9	233	1	YOD2_CABEL	P34594 caenorhabdi
27	7	0.9	249	1	PSB1_HUMAN	Q06323 h proteasom
28	7	0.9	249	1	PSE1_WACFA	P58238 macaca fasc
29	7	0.9	251	1	1433_FUCVE	Q39757 fucus vesic
30	7	0.9	251	1	A32B_HUMAN	Q92688 homo sapien
31	7	0.9	254	1	OAZ_DROME	P54361 drosophila
32	7	0.9	256	1	YP96_CABEL	Q09476 caenorhabdi
33	7	0.9	264	1	KSGA_FUSNN	Q8r6b1 f dimethyla

34 7 0.9 268 1 KSGA_PSEAE Q915u5 p dimethyla
35 7 0.9 268 1 TRPA_PASMU P57854 pasteurella
36 7 0.9 272 1 APHE_STRGR P18150 streptomyce
37 7 0.9 301 1 SECU_SCHPO P21135 schizosacch
38 7 0.9 328 1 P2Y6_HUMAN Q15077 homo sapien
39 7 0.9 328 1 P2Y6_RAT Q63371 rattus norv
40 7 0.9 338 1 LEGB_PEA P14594 pisum sativ
41 7 0.9 347 1 CCA_SCHMA O02197 schistosoma
42 7 0.9 349 1 YHHT_ECOLI P37622 escherichia
43 7 0.9 350 1 SYFA_LISIN Q92ci7 listeria in
44 7 0.9 350 1 SYFA_LISMO Q8y7q2 listeria mo
45 7 0.9 357 1 SYFA_CAUCR Q9a9e4 caulobacter

ALIGNMENTS

RESULT 1

EPLI_HUMAN
ID AC Q9UHB6; Q9BVF2; Q9H8J1; Q9HBN5; Q9NX96; Q9NXC3; Q9NXU6; Q9POH8;
AC Q9UHB5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Epithelial protein lost in neoplasm.
GN EPLIN OR SREBP3.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
RC TISSUE=Cervical carcinoma;
RX MEDLINE=20087188; PubMed=10618726;
RA Maul R.S., Chang D.D.;
RT "EPLIN, epithelial protein lost in neoplasm."
RL Oncogene 18:7838-7841(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
RX MEDLINE=20267849; PubMed=10806352;
RA Chen S., Maul R.S., Kim H.R., Chang D.D.;
RT "Characterization of the human EPLIN (Epithelial protein lost in neoplasm) gene reveals distinct promoters for the two EPLIN isoforms."
RL Gene 248:69-76(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM BETA).
RT TISSUE=Uterus;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S., Ansoerge W., Boecher M., Bloeker H., Bauersachs S., Blum H., Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N., Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D., Wambutt R., Korn B., Klein M., Boustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs."
RL Genome Res. 11:422-435(2001).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA; BETA AND 3).
RC TISSUE=Colon, Hepatoma, and Placenta;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Hypothalamus;
RX MEDLINE=20402571; PubMed=10931946;
RA Hu R.-M., Han Z.-G., Song H.-D., Peng Y.-D., Huang Q.-H., Ren S.-X., Gu Y.-J., Huang C.-H., Li Y.-B., Jiang C.-L., Fu G., Zhang Q.-H., Gu B.-W., Dai M., Mao Y.-F., Gao G.-F., Rong R., Ye M., Zhou J.,

RA Xu S.-H., Gu J., Shi J.-X., Jin W.-R., Zhang C.-K., Wu T.-M.,
RA Huang G.-Y., Chen Z., Chen M.-D., Chen J.-L.,
RT "Gene expression profiling in the human hypothalamus-pituitary-adrenal
RT axis and full-length cDNA cloning.",
RL Proc. Natl. Acad. Sci. U.S.A. 97:9543-9548(2000).
RN [6]
RN SEQUENCE OF 232-759 FROM N.A.
RP Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.O., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.,
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.",
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RN SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC TISSUE:Colon, and Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bustow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loguercio N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettunen M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalius D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.",
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. This cytoskeletal protein co-
CC localizes with actin stress fibers and focal adhesion plaques.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=Beta;
CC IsoId=Q9UHB6-1; Sequence=Displayed;
CC Name=Alpha;
CC IsoId=Q9UHB6-2; Sequence=VSP_003116;
CC Name=3;
CC IsoId=Q9UHB6-3; Sequence=VSP_003117;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta, kidney,
CC pancreas, prostate, ovary, spleen and heart. Also detected in
CC lung, liver, brain, skeletal muscle, thymus, testis and intestine.
CC Not detected in leukocytes. Epln-beta levels are generally very
CC low. Epln-alpha is abundant in epithelial cells from mammary
CC gland, prostate and in normal oral keratinocytes. Levels are low
CC in aortic endothelial cells and dermal fibroblasts, and not
CC detectable in myocardium.
CC -!- INDUCTION: Down-regulated in some cancer cell lines. Epln-alpha
CC is induced by serum. Epln-beta is constitutively expressed.
CC -!- SIMILARITY: Contains 1 LIM zinc-binding domain.
CC -!- CAUTION: Ref.4 (BAA91120) sequence differs from that shown due to
CC a frameshift in position 697.
CC -!- CAUTION: Ref.6 sequence differs from that shown due to frameshifts
CC in positions 365 and 662.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; AF198454; AAF23755.1; -
DR EMBL; AF198455; AAF23756.1; -
DR EMBL; AF136911; CAB66845.1; -
DR EMBL; AK000372; BAA91120.1; ALT_FRAME.
DR EMBL; AK000335; BAA91092.1; -
DR EMBL; AK003649; BAB14625.1; -
DR EMBL; AK000057; BAA90914.1; -
DR EMBL; AF157325; AAF67491.1; -
DR EMBL; AF18025; AAG17267.1; ALT_FRAME.
DR EMBL; BC001247; AAH01247.1; -
DR EMBL; BC010664; AAH00664.1; -
DR HSP; P32965; ICTL.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS00023; LIM_DOMAIN_2; 1.
KW Cytoskeleton; LIM domain; Metal-binding; Zinc; Alternative splicing.
FT DOMAIN 388 448 LIM.
FT VARSPLIC 1 160 Missing (in isoform Alpha).
FT VARSPLIC 1 302 Missing (in isoform 3).
FT CONFLICT 344 344 R -> PG (IN REF. 1; AAF23756).
FT CONFLICT 381 381 MISSING (IN REF. 4; BAA90914 AND 7).
FT CONFLICT 415 415 F -> L (IN REF. 6).
FT CONFLICT 463 463 D -> G (IN REF. 4; BAA90914).
FT CONFLICT 491 491 P -> Q (IN REF. 5).
FT CONFLICT 520 521 DK -> NR (IN REF. 5).
SQ SEQUENCE 759 AA; 85225 MW; 996378AFD3B003D5 CRC64;
Query Match 100.0%; Score 759; DB 1; Length 759;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MESSPFNRQWTSLSLRVTAKLSLVNKKSAIIVEIFSKYKAAAEETNMKKRSNTNL 60
DB 1 MESSPFNRQWTSLSLRVTAKLSLVNKKSAIIVEIFSKYKAAAEETNMKKRSNTNL 60
QY 61 SQHFRKGTTLVLKKWENPGLCAESHTSLRNSSTTEIRHRADHPAETVSHAASGAKADQ 120
DB 61 SQHFRKGTTLVLKKWENPGLCAESHTSLRNSSTTEIRHRADHPAETVSHAASGAKADQ 120
QY 121 EQIHPRSRLRSPPEALVQGRYPHIKDGEDLKDHSSTESKMKMNCCLGESRHEVEKSEISEN 180
DB 121 EQIHPRSRLRSPPEALVQGRYPHIKDGEDLKDHSSTESKMKMNCCLGESRHEVEKSEISEN 180
QY 181 TDASGKIEKYNVPLNRLKXWPEKGEPTQTTLRAOSRSASGRKISSENSYSLDLEIGPGQ 240
DB 181 TDASGKIEKYNVPLNRLKXWPEKGEPTQTTLRAOSRSASGRKISSENSYSLDLEIGPGQ 240
QY 241 LSSSTFDSEKSNRRNLELPRLSETSIKDRMAKYAAVSKOSSSTNYTNELKASGEIKI 300
DB 241 LSSSTFDSEKSNRRNLELPRLSETSIKDRMAKYAAVSKOSSSTNYTNELKASGEIKI 300
QY 301 HKWEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVSKSEVQVHPKP 360
DB 301 HKWEQKENVPPGPEVCITHQEGEKISANENSLAVRSTPAEDDSRDSQVSKSEVQVHPKP 360
QY 361 LSPDRASLSSESSPPKAMKXFPAPARETCVCOQTVYPMERLLANQQVFIHSCFRSYC 420
DB 361 LSPDRASLSSESSPPKAMKXFPAPARETCVCOQTVYPMERLLANQQVFIHSCFRSYC 420
QY 421 NNKLSLGYASLHRIYCKPHNQLFKSGKGYDGFGRPHKDLWASKNNEEILERPQAQ 480
DB 421 NNKLSLGYASLHRIYCKPHNQLFKSGKGYDGFGRPHKDLWASKNNEEILERPQAQ 480
QY 481 LANARETPHSGVEDAPIAKVGVLAASMEAKASQOEKEDKPAETKKLRITAWPPPTLGS 540
DB 481 LANARETPHSGVEDAPIAKVGVLAASMEAKASQOEKEDKPAETKKLRITAWPPPTLGS 540

KW Metal-binding; Zinc; Transcription regulation; Activator.
FT DOMAIN 28 78 LIM 1.
FT DOMAIN 87 141 LIM 2.
FT DNA_BIND 154 213 HOMEOBOX.
SQ SEQUENCE 398 AA; 44226 MW; F0298F0D1A725C1F CRC64;

Query Match 1.1%; Score 8; DB 1; Length 398;
Best Local Similarity 100.0%; Pred.No.9.8; 0; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

QY 368 SSLSESSP 375
| | | | | | | |
Db 272 SSLSESSP 279

RESULT 4
ID YS63 CAEEL STANDARD; PRT; 405 AA.
AC Q09179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK675.3 in chromosome II.
GN ZK675.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Sims M.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BESTOPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z46912; CAA86845.1; -;
DR PIR; T27971;
DR WormPep; ZK675.3; CE01722.
DR InterPro; IPR000615; Worm_fam_8.
DR Pfam; PF01062; DUF289; 1.
DR ProDom; PD002802; Worm_fam_8; 1.
KW Hypothetical protein.
SQ SEQUENCE 405 AA; 47835 MW; 0F589D874E9E0E61 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 405;
Best Local Similarity 100.0%; Pred.No.10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 SKNEEEI 474
| | | | | | | |
Db 254 SKNEEEI 261

RESULT 5
MTDI DESDN STANDARD; PRT; 415 AA.
AC P05302;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase DdeI (EC 2.1.1.73) (Cytosine-specific
DE methyltransferase DdeI) (M.DdeI).
GN DDEIM OR DDEM.
OS Desulfovibrio desulfuricans (strain Norway 4).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;

OC Desulfomicrobiaceae; Desulfomicrobium.
OX NCBI_TaxID=52561;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88040453; PubMed=2823226;
RA Snyder L.A., Slatko B., Moran L., O'Donnell K.H., Brooks J.E.;
RT "Nucleotide sequence of the DdeI restriction-modification system and
RT characterization of the methylase protein.";
RL Nucleic Acids Res. 15:8249-8266(1987)
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC CTNAG. CAUSES SPECIFIC METHYLATION ON C-1 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE DDEI ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA cytosine = S-
CC adenosyl-L-homocysteine + DNA 5-methylcytosine.
CC -!- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y00449; CAA68505.1; -;
DR HSSP; P20589; 1DCT
DR REBASE; 3357; M.DdeI.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF00145; DNA_methylase; 1.
DR PRINTS; PRO0105; C5METTRFRASE.
DR TIGRFAMs; TIGR00675; dcm; 1.
DR PROSITE; PS00094; C5_MTASE_1; 1.
DR PROSITE; PS00095; C5_MTASE_2; 1.
KW Transferase; Methyltransferase; Restriction system.
FT ACT_SITE 76 76 BY SIMILARITY.
SQ SEQUENCE 415 AA; 47081 MW; 34DF5968E0414EED CRC64;

Query Match 1.1%; Score 8; DB 1; Length 415;
Best Local Similarity 100.0%; Pred.No.10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 TVASFSQS 599
| | | | | | | |
Db 296 TVASFSQS 303

RESULT 6
SELA RHIME
ID SELA RHIME STANDARD; PRT; 466 AA.
AC P58226;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE L-seryl-tRNA(Sec) selenium transferase (EC 2.9.1.1) (Cysteiny-
DE tRNA(Sec) selenium transferase) (Selenocysteine synthase)
DE (Selenocysteinyl-tRNA(Sec) synthase).
GN SELA OR RA0005 OR SMA0011.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kaiman S., Keating D.H., Palm C., Beck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";

```

RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -!- FUNCTION: Converts seryl-tRNA(Sec UCA) to selenocysteinyl-tRNA(Sec
CC UCA) during selenoprotein biosynthesis (By similarity).
CC -!- CATALYTIC ACTIVITY: L-seryl-tRNA(Sec) + selenophosphate = L-
CC selenocysteinyl-tRNA(Sec) + H2O + phosphate.
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SELA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE007195; AAK64663.1; -
DR PIR; E95262; E95262.
DR HAMAP; MF_00423; - 1.
DR InterPro; IPR004534; Sela.
DR Pfam; PF03841; Sela; 1.
DR TIGRFAMs; TIGR00474; sela; 1.
DR Transferase; Pyridoxal phosphate; Selenium; Plasmid;
KW Complete proteome.
FT BINDING 292 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 466 AA; 49028 MW; B20B8852451C2EAD CRC64;

Query Match 1.1%; Score 8; DB 1; Length 466;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 GSSGSALE 546
DB 411 GSSGSALE 418
|||||

RESULT 7
LIK2_CHICK ID LIK2_CHICK STANDARD; PRT; 642 AA.
AC P53666;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE LIM domain kinase 2 (EC 2.7.1.-) (LIK2).
GN LIK2 OR LIK2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=9515277; PubMed=7852284;
RA Ohashi K., Toshima J., Tajima K., Nakamura T., Mizuno K.;
RT "Molecular cloning of a chicken lung cDNA encoding a novel protein
RL J. Biochem. 116:636-642(1994).
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE LUNG, AND
CC FAINTLY IN THE KIDNEY, LIVER, BRAIN, SPLEEN, GIZZARD, AND
CC INTESTINE.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -!- SIMILARITY: Contains 2 LIM zinc-binding domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; D26310; BAA05372.1; -
DR PIR; JP0079; JP0079.
DR HSP; P32965; 1CTL.
DR InterPro; IPR001781; LIM.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000094; LIM; 2.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00132; LIM; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 2.
DR PROSITE; PS00023; LIM_DOMAIN_2; 2.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat;
KW LIM domain; Metal-binding; Zinc.
FT DOMAIN 12 63 LIM 1.
FT DOMAIN 72 124 LIM 2.
FT DOMAIN 152 239 PDZ.
FT DOMAIN 331 608 PROTEIN_KINASE.
FT NP_BIND 337 345 ATP (BY SIMILARITY).
FT BINDING 360 360 ATP (BY SIMILARITY).
FT ACT_SITE 451 451 BY SIMILARITY.
SQ SEQUENCE 642 AA; 72467 MW; 6D3E500B8C3A8277 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 642;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 HISCFRCS 418
DB 34 HISCFRCS 41
|||||

RESULT 8
GARB_AQUAE ID GARB_AQUAE STANDARD; PRT; 792 AA.
AC O67137;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA gyrase subunit B (EC 5.99.1.3).
GN GYRB OR AQ_1026.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=VFS;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujaay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RL Nature 392:353-358(1998).
CC -!- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE

```



```
CC ENZYME FORMS AN A2B2 TETRAMER.
CC -I- SIMILARITY: Belongs to the type II topoisomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000719; AAC07098.1; -
CC DR PIR; F70388; F70388.
CC DR HSPG; P06982; 1AJ6.
CC DR InterPro; IPR003594; ATPbind_ATPase.
CC DR InterPro; IPR002288; DNA_gyrase_C.
CC DR InterPro; IPR000585; DNA_gyrase.
CC DR InterPro; IPR001241; DNA_topoisomII.
CC DR InterPro; IPR006171; Toprim_dom.
CC DR Pfam; PF00204; DNA_gyrase; 1.
CC DR Pfam; PF00986; DNA_gyraseB; 1.
CC DR Pfam; PF02518; HATPase_C; 1.
CC DR Pfam; PF01751; Toprim; 1.
CC DR PRINTS; PR00418; TP12FAMILY.
CC DR ProDom; PD149633; DNA_gyraseB_C; 1.
CC DR SMART; SM00387; HATPase_C; 1.
CC DR SMART; SM00433; TOP2C; 1.
CC DR TIGRams; TIGR01059; gyrB; 1.
CC DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
CC DR Topoisomerase; Isomerase; ATP-binding; Complete proteome.
CC KW SEQUENCE 792 AA; 90530 MW; 07DC02DF293FF9F3 CRC64;
SQ -----
Query Match 1.1%; Score 8; DB 1; Length 792;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 572 DLDLKKLR 579
Db 486 DLDLKKLR 493
|||||
RESULT 9
BMR2_HUMAN STANDARD; PRT; 1038 AA.
AC Q1873; Q18569;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bone morphogenetic protein receptor type II precursor (EC 2.7.1.37)
DE (BMP type II receptor) (BMPR-II).
GN BMPR2 OR PPH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Substantia nigra;
RX MEDLINE=95372334; PubMed=7644468;
RA Rosenzweig B.L., Imamura T., Okadome T., Cox G.N., Yamashita H.,
RA ten Dijke P., Heldin C., Miyazono K.;
RT "Cloning and characterization of a human type II receptor for bone
RT morphogenetic proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 92:7632-7636(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin fibroblast;
RX MEDLINE=98403457; PubMed=7673243;
RA Nohno T., Ishikawa T., Saito T., Hosokawa K., Noji S., Wosing D.H.,
RA Rosenbaum J.S.;
RT "Identification of a human type II receptor for bone morphogenetic
RT protein-4 that forms differential heteromeric complexes with bone
RT morphogenetic protein type I receptors."
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95197572; PubMed=7890683;
RA Kawabata M., Chytil A., Moses H.L.;
RT "Cloning of a novel type II serine/threonine kinase receptor through
RT interaction with the type I transforming growth factor-beta
RT receptor."
RL J. Biol. Chem. 270:5625-5630(1995).
RN [4]
RP VARIANTS PPH1 GLN-491 AND TRP-491.
RX MEDLINE=20395844; PubMed=10903931;
RA Deng Z., Morse J.H., Slager S.L., Cuervo N., Moore K.J., Venetos G.,
RA Kalachikov S., Cayanis E., Fischer S.G., Barst R.J., Hodge S.E.,
RA Knowles J.A.;
RT "Familial primary pulmonary hypertension (gene PPH1) is caused by
RT mutations in the bone morphogenetic protein receptor-II gene."
RL Am. J. Hum. Genet. 67:737-744 (2000).
RN [5]
RP VARIANTS PPH1 TYR-60; TYR-117 AND ARG-483.
RX MEDLINE=20473811; PubMed=11015450;
RA Thomson J.R., Machado R.D., Paucululo M.W., Morgan N.V., Humbert M.,
RA Elliott G.C., Ward K., Yacoub M., Mikhail G., Rogers P., Newman J.H.,
RA Wheeler L., Higenbottam T., Gibbs J.S.R., Egan J., Crozier A.,
RA Peacock A., Allcock R., Corris P., Loyd J.E., Trembath R.C.,
RA Nichols W.C.;
RT "Sporadic primary pulmonary hypertension is associated with germline
RT mutations of the gene encoding BMPR-II, a receptor member of the
RT TGF-beta family."
RL J. Med. Genet. 37:741-745 (2000).
RN [6]
RP VARIANTS PPH1 TRP-118; TYR-347 AND GLY-485.
RX MEDLINE=20428187; PubMed=10973254;
RA Lane K.B., Machado R.D., Paucululo M.W., Thomson J.R.,
RA Phillips J.A. III, Loyd J.E., Nichols W.C., Trembath R.C., Aldred M.,
RA Brannon C.A., Conneally P.M., Foroud T., Fretwell N., Gaddipati R.,
RA Koller D., Loyd E.J., Morgan N.V., Newman J.H., Gaddipati R.,
RA Villano Gueell C., Wheeler L.;
RT "Heterozygous germline mutations in BMPR2, encoding a TGF-beta
RT receptor, cause familial primary pulmonary hypertension."
RL Nat. Genet. 26:81-84(2000).
RN [7]
RP VARIANTS PPH1 ARG-123; SER-123; ARG-420 AND THR-512, VARIANT ASP-224,
RX AND CHARACTERIZATION OF VARIANT PPH1 GLY-485.
RX MEDLINE=21063176; PubMed=11115378;
RA Machado R.D., Paucululo M.W., Thomson J.R., Lane K.B., Morgan N.V.,
RA Wheeler L., Phillips J.A. III, Newman J.H., Williams D., Galie N.,
RA Manes A., McNeil K., Yacoub M., Mikhail G., Rogers P., Corris P.,
RA Humbert M., Donnai D., Martensson G., Tranebjærg L., Loyd J.E.,
RA Trembath R.C., Nichols W.C.;
RT "BMPR2 haploinsufficiency as the inherited molecular mechanism for
RT primary pulmonary hypertension."
RL Am. J. Hum. Genet. 68:92-102(2001).
CC -I- FUNCTION: BINDS TO BMP-7, BMP-2 AND, LESS EFFICIENTLY, BMP-4.
CC BINDING IS WEAK BUT ENHANCED BY THE PRESENCE OF TYPE I RECEPTORS
CC FOR BMPs.
CC -I- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -I- SUBUNIT: HETERODIMERIZE WITH TYPE-I RECEPTORS.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART AND LIVER.
CC -I- DISEASE: Defects in BMPR2 are the cause of primary pulmonary
CC hypertension (PPH1) [MIM:178600]; a rare autosomal dominant
CC disorder characterized by plexiiform lesions of proliferating
CC endothelial cells in pulmonary arterioles. The lesions lead to
CC elevated pulmonary arterial pressure, right ventricular failure,
CC and death. The disease can occur from infancy throughout life and
CC it has a mean age at onset of 36 years. Penetrance is reduced.
CC Although familial PPH1 is rare, cases secondary to known
CC etiologies are more common and include those associated with the
CC appetite-suppressant drugs.
CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC TGF-beta RECEPTOR SUBFAMILY.
CC -----
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

-----
DR EMBL; Z48923; CAAB8759.1; -.
DR EMBL; D50516; BAA09094.1; -.
DR EMBL; U20165; AAC50105.1; -.
DR PIR; J38935; J38935..
DR Genew; HGNC:1078; BMPR2..
DR MIN; 178600; -.
DR GO; GO:0005887; C: integral to plasma membrane; TAS.
DR GO; GO:0005515; F: protein binding activity; TAS.
DR GO; GO:0007118; P: transmembrane receptor protein serine/threonine/threo. .; TAS.
DR InterPro; IPR000472; Activin_rec.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_Thr_pkinase.
DR Pfam; PF01064; Activin_rec; 1.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; FALSE_NEG.
DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.
DR Receptor; Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Transmembrane, glycoprotein; Signal; Polymorphism; Disease mutation.
FT SIGNAL 1 26
FT CHAIN 27 1038
FT DOMAIN 27 150 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 151 171 POTENTIAL.
FT DOMAIN 172 1038 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 203 504 PROTEIN KINASE.
FT NP_BIND 209 217 ATP (BY SIMILARITY).
FT BINDING 230 230 ATP (BY SIMILARITY).
FT ACT_SITE 333 333 BY SIMILARITY.
FT DOMAIN 547 550 POLY-SER.
FT DOMAIN 610 618 POLY-THR.
FT DOMAIN 901 908 POLY-ASN.
FT CARBOHYD 55 55 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 60 60 C -> Y (in PPH1).
FT /FTID=VAR_013670.
FT VARIANT 117 117 /FTID=VAR_013671.
FT VARIANT 118 118 C -> W (in PPH1).
FT /FTID=VAR_013672.
FT VARIANT 123 123 C -> R (in PPH1).
FT /FTID=VAR_013673.
FT VARIANT 123 123 C -> S (in PPH1).
FT /FTID=VAR_013674.
FT VARIANT 224 224 E -> D.
FT /FTID=VAR_013675.
FT VARIANT 347 347 C -> Y (in PPH1).
FT /FTID=VAR_013676.
FT VARIANT 420 420 C -> R (in PPH1).
FT /FTID=VAR_013677.
FT VARIANT 483 483 C -> R (in PPH1; sporadic).
FT /FTID=VAR_013678.
FT VARIANT 485 D -> G (in PPH1; complete loss of function).
FT /FTID=VAR_013679.
FT VARIANT 491 R -> Q (in PPH1; sporadic).
FT /FTID=VAR_013680.
FT VARIANT 491 R -> W (in PPH1).
FT /FTID=VAR_013681.
FT VARIANT 512 K -> T (in PPH1).
FT /FTID=VAR_013682.
FT VARIANT 519 N -> K (in PPH1).

```

```

FT CONFLICT 828 828 /FTID=VAR_013683.
FT G -> R (in REF. 1).
SQ SEQUENCE 1038 AA; 115201 MW; 1389923CE574B913 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 1038;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 ENSLAVRS 336
Db 468 ENSLAVRS 475
|||||

RESULT 10
BMR2_MOUSE
ID BMR2_MOUSE STANDARD; PRT; 1038 AA.
AC Q35607;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bone morphogenetic protein receptor type II precursor (EC 2.7.1.37)
DE (BMP type II receptor) (BMPRII) (BRK-3).
DE BMPR2.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97350808; PubMed=9207184;
RA Beppu H., Minowa O., Miyazono K., Kawabata M.;
RT "cDNA cloning and genomic organization of the mouse BMP type II
RT receptor.";
RL Biochem. Biophys. Res. Commun. 235:499-504 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Whitaker G.B., Koenig B.B., Ting J., Tiesman J.P., Limberg A.L.,
RA Grant R.A., Begley K.B., Rosenbaum J.S.;
RT "Identification of BMP receptor complexes with differential signaling
RT properties and ligand binding profiles.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS TO BMP-7, BMP-2 AND, LESS EFFICIENTLY, BMP-4.
CC BINDING IS WEAK BUT ENHANCED BY THE PRESENCE OF TYPE I RECEPTORS
CC FOR BMPs.
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- SUBUNIT: HETERODIMERIZE WITH TYPE-I RECEPTORS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC TGF RECEPTOR SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

KW Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 26
FT CHAIN 27 1038
FT
FT DOMAIN 27 150
FT TRANSMEM 151 171
FT DOMAIN 172 1038
FT DOMAIN 203 504
FT NP_BIND 209 217
FT BINDING 230 230
FT ACT_SITE 333 333
FT DOMAIN 191 194
FT DOMAIN 547 550
FT DOMAIN 610 618
FT DOMAIN 901 908
FT CARBOHYD 55 55
FT CARBOHYD 110 110
FT CARBOHYD 126 126
FT CARBOHYD 126 126
SQ SEQUENCE 1038 AA; 115019 MW; 4106945DC63250E1 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 1038;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 ENSLAVERS 336
DB 468 ENSLAVERS 475

RESULT 11
YGFY_HAEIN STANDARD; PRT; 85 AA.
AC P44025.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein HI0627.
GN HI0627.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu D.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
RT Science 269:496-512(1995).
RL
CC -1- SIMILARITY: STRONG, TO E.COLI YGFY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32745; AAC22287.1;
CC PIR; G64010; G64010.
CC TIGR; HI0627;
CC InterPro; IPR005631; DUF339.
CC Pfam; PF03937; TPR_div1; 1.

```

```

KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 85 AA; 10430 MW; AAE707492162A1C CRC64;

Query Match 0.9%; Score 7; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 KSQDEL 731
DB 63 KSQDEL 69

RESULT 12
SUI1_AERPE STANDARD; PRT; 101 AA.
AC Q5YBG9.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein translation factor SUI1 homolog.
GN APE1629.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takehashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, Aeropyrum pernix K1."
RT DNA Res. 6:83-101(1999).
RL
CC -1- SIMILARITY: BELONGS TO THE SUI1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000062; BAA80629.1;
CC PIR; H72542; H72542.
CC HSSP; F08245; 1DIR.
CC HAMAP; MF_00604; -; 1.
CC InterPro; IPR005872; SUI1_arch_bact.
CC InterPro; IPR001950; TIF_SUI1.
CC Pfam; PF01253; SUI1; 1.
CC TIGRFAMs; TIGR01158; SUI1_rel; 1.
CC PROSITE; PS01118; SUI1_1; 1.
CC PROSITE; PS02096; SUI1_2; 1.
CC Protein biosynthesis; Translation regulation; Complete proteome.
SQ SEQUENCE 101 AA; 11470 MW; 80F377A4B658D636 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 740 LSVEEQI 746
DB 18 LSVEEQI 24

RESULT 13
HMGL_VICFA
ID "HMGL_VICFA STANDARD; PRT; 149 AA.

```

P40620;
 01-FEB-1995 (Rel. 31, Created)
 01-FEB-1995 (Rel. 31, Last sequence update)
 15-SEP-2003 (Rel. 42, Last annotation update)
 HMGI/2-like protein.
 Vicia faba (Broad bean).
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Vicia.
 NCBI_TaxID=3906;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=cv. Minor; TISSUE=Cotyledon;
 MEDLINE=9403341; PubMed=8219095;
 Grasser K., Wohlfarth T., Baumeister H., Felix G.;
 "Comparative analysis of chromosomal HMG proteins from monocotyledons
 and dicotyledons";
 Plant Mol. Biol. 23:619-625(1993).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: BELONGS TO THE HMGI/HMG2 PROTEIN FAMILY.
 CC -!- SIMILARITY: Contains 1 HMG box domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 EMBL; Z21703; CAB37859.1; -;
 PIR; S39556; S39556.
 DR HSSP; P07155; IHMF.
 DR InterPro; IPR000910; HMG_12_box.
 DR Pfam; PF00505; HMG_box; 1.
 DR SMART; SM00398; HMG; 1.
 DR PROSITE; PS0118; HMG_BOX_2; 1.
 DR Nuclear protein; DNA-Binding.
 KW DNA_BIND 45 114 HMG_BOX.
 FT DOMAIN 121 149 ASP/GLU-RICH (ACIDIC).
 FT SEQUENCE 149 AA; 16659 MW; FE52DF5E68E0B3D CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 753 DEDEEE 759
 Db 135 DEDEEE 141
 [1]
 RESULT 14
 ID PYRI METAC STANDARD; PRT; 156 AA.
 AC Q8THL3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Aspartate carbamoyltransferase regulatory chain.
 GN PYRI OR MA4501.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 Linton L., McEwan P., McKernan K., Talanas J., Tirrell A., Ye W.,
 Zimmer A., Barber R.D., Cann I., Graham D.E., Guss A.M.,

Hedderich R., Ingram-Smith C., Kuettner H.C., Krzyzski J.A.,
 Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 Metcalf W.W., Birren B.;
 "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity";
 Genome Res. 12:532-542(2002).
 CC -!- FUNCTION: Involved in allosteric regulation of aspartate
 carbamoyltransferase (By similarity).
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SUBUNIT: Contains catalytic and regulatory chains.
 CC -!- SIMILARITY: BELONGS TO THE PYRI FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 EMBL; AE011170; AM07841.1; -;
 HAMAp; MF_00002; -; 1.
 DR InterPro; IPR002801; Pyri.
 DR Pfam; PF01948; Pyri; 1.
 DR Pfam; PF02748; Pyri_C; 1.
 DR ProDom; PD006194; Pyri; 1.
 DR TIGRFAMs; TIGR00240; Atcasc_reg; 1.
 KW Pyrimidine biosynthesis; Zinc; Complete proteome.
 FT METAL 109 109 ZINC (BY SIMILARITY).
 FT METAL 114 114 ZINC (BY SIMILARITY).
 FT METAL 140 140 ZINC (BY SIMILARITY).
 FT METAL 143 143 ZINC (BY SIMILARITY).
 SQ SEQUENCE 156 AA; 17336 MW; 51DF22654D3B83A1 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 156;
 Best Local Similarity 100.0%; Pred No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 738 KELSVEE 744
 Db 64 KELSVEE 70
 [1]
 RESULT 15
 ID PYRI METWA STANDARD; PRT; 156 AA.
 AC Q8PXK6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Aspartate carbamoyltransferase regulatory chain.
 GN PYRI OR MW1212.
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=Goel / GoI / ATCC BAA-199 / DSM 3647 / OCM 88;
 RX MEDLINE=22120827; PubMed=12125824;
 RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
 Martinez-Arias R., Henne A., Wietzer A., Baumer S., Jacobi C.,
 Bruggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
 Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
 Fritz H.-J., Gottschalk G.;
 "The genome of Methanosarcina mazei: evidence for lateral gene
 transfer between Bacteria and Archaea";
 J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 CC -!- FUNCTION: Involved in allosteric regulation of aspartate
 carbamoyltransferase (By similarity).
 CC

```
CC -!- SUBUNIT: Contains catalytic and regulatory chains.
CC -!- SIMILARITY: BELONGS TO THE PYRI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE013351; AAM30908.1; -
DR HAMAP; MF 00002; -; 1.
DR InterPro; IPR002801; Pyri.
DR Pfam; PF01948; Pyri.1.
DR Pfam; PF02748; Pyri.C; 1.
DR PRODOM; PD006194; Pyri; 1.
DR TIGRFAMs; TIGR00240; AtCase; reg; 1.
KW Pyrimidine biosynthesis; Zinc; Complete proteome.
FT METAL 109 109 ZINC (BY SIMILARITY).
FT METAL 114 114 ZINC (BY SIMILARITY).
FT METAL 140 140 ZINC (BY SIMILARITY).
FT METAL 143 143 ZINC (BY SIMILARITY).
SQ SEQUENCE 156 AA; 17011 MW; B77F6C13F57BA962 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 156;
Best Local Similarity 100.0%; Pred.No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 738 KELSVEE 744
Db 64 KELSVEE 70
|||||
```

Search completed: January 6, 2004, 09:52:13
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:49:05 ; Search time 41 Seconds
(without alignments)
4777.120 Million cell updates/sec

Title: US-09-890-549-4

Perfect score: 759

Sequence: 1 MESSFPNRRQWTSLSLRVTA.....LSVEEQIKRNYDEDEEE 759

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 810525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 810525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	9.1	593	11 Q8CD09	Q8cd09 mus musculus
2	69	9.1	593	11 Q8BT15	Q8bt15 mus musculus
3	69	9.1	753	11 Q8X2H0	Q8x2h0 mus musculus
4	69	9.1	753	11 Q8C7S2	Q8c7s2 mus musculus
5	65	8.6	593	11 Q8C3R7	Q8c3r7 mus musculus
6	15	2.0	519	4 Q8N7Z0	Q8n7z0 homo sapien
7	15	2.0	951	6 Q8HXJ9	Q8hxj9 macaca fasc
8	14	1.8	127	4 Q9BT23	Q9bt23 homo sapien
9	14	1.8	127	4 Q9BS91	Q9bs91 homo sapien
10	14	1.8	128	11 Q8BGB5	Q8bgb5 mus musculus
11	14	1.8	629	13 Q9DEY8	Q9dey8 brachydanio
12	9	1.2	647	12 Q9DWB9	Q9dwb9 rat cytonem
13	8	1.1	101	11 Q62493	Q62493 mus musculus
14	8	1.1	120	2 Q9ALE6	Q9ale6 uncultured
15	8	1.1	122	17 Q97Z14	Q97z14 sulfobolus
16	8	1.1	135	6 Q8HXR3	Q8hxr3 bos taurus

17	8	1.1	148	9 Q9B095	Q9b095 mycobacteri
18	8	1.1	169	17 Q8TV97	Q8tv97 methanopyru
19	8	1.1	231	10 Q8O964	Q8o964 arabisdopsis
20	8	1.1	234	17 Q9YG41	Q9yg41 aeropyrum p
21	8	1.1	279	5 Q9V5Z4	Q9v5z4 drosophila
22	8	1.1	297	4 Q9A5A1	Q9a5a1 homo sapien
23	8	1.1	354	5 Q76Z76	Q76z76 drosophila
24	8	1.1	358	4 Q96BE7	Q96be7 homo sapien
25	8	1.1	365	11 Q8BHK0	Q8bhk0 mus musculu
26	8	1.1	374	13 Q42N86	Q42n86 brachydanio
27	8	1.1	386	10 Q9FN10	Q9fn10 arabisdopsis
28	8	1.1	388	17 Q9EY10	Q9ey10 aeropyrum p
29	8	1.1	403	10 Q42381	Q42381 arabisdopsis
30	8	1.1	413	10 Q64815	Q64815 arabisdopsis
31	8	1.1	445	16 Q99YL0	Q99yl0 streptococc
32	8	1.1	445	16 Q8NZY9	Q8nzy9 streptococc
33	8	1.1	445	16 Q8K6D5	Q8ked5 streptococc
34	8	1.1	459	5 Q9VCB8	Q9vcb8 drosophila
35	8	1.1	459	5 Q9LV69	Q9lv69 arabisdopsis
36	8	1.1	473	16 Q8DLI3	Q8dl13 synechococc
37	8	1.1	530	4 Q13161	Q13161 homo sapien
38	8	1.1	682	10 Q94GH0	Q94gh0 oryza sativ
39	8	1.1	730	5 Q9VC06	Q9vc06 drosophila
40	8	1.1	784	16 Q9HXB2	Q9hxb2 pseudomonas
41	8	1.1	829	10 Q8LIY9	Q8liy9 oryza sativ
42	8	1.1	874	5 Q95X65	Q95x65 caenorhabdi
43	8	1.1	942	5 Q95X66	Q95x66 caenorhabdi
44	8	1.1	1026	10 Q9SS79	Q9ss79 arabisdopsis
45	8	1.1	1031	13 Q42124	Q42124 gallus gall

ALIGNMENTS

RESULT 1

Q8CD09 Q8CD09 PRELIMINARY; PRT; 593 AA.
AC Q8CD09;

DT 01-MAR-2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Epithelial protein lost in neoplasm.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RX MEDLINE=22354683; PubMed=12468851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RL 60,770 full-length cDNAs."

RL Nature 420:563-573(2002).

DR EMBL; AK031698; BAC27520.1;

SQ SEQUENCE 593 AA; 66017 MW; 122AB8AA07E1FCOA CRC64;

Query Match

Best Local Similarity 9.1%; Score 69; DB 11; Length 593;

Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 CVEQKTVYMERLLANQQVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHNQLPKSK 449

DB 228 CVEQKTVYMERLLANQQVFHISCFRCSCYNNKLSLGTYSALHGRYCKPHNQLPKSK 287

QY 450 GNYDEGFCH 458

DB 288 GNYDEGFCH 296

RESULT 2

Q8BT15

```
ID Q8BT15 PRELIMINARY; PRT; 593 AA.
AC Q8BT15;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK028186; BAC25798.1; -.
SQ SEQUENCE 593 AA; 66026 MW; 3F2AB8AA07F7BE6E CRC64;

Query Match 9.1%; Score 69; DB 11; Length 593;
Best Local Similarity 100.0%; Pred. No. 1.3e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 CVECQKTVPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYICKPHFNQLFKSK 449
DB 228 CVECQKTVPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYICKPHFNQLFKSK 287
QY 450 GNYDEGFGH 458
DB 288 GNYDEGFGH 296

RESULT 3
Q8K2H0 PRELIMINARY; PRT; 753 AA.
ID Q8K2H0;
AC Q8K2H0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RA Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
DR EMBL; BC031490; AAH31490.1; -.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SMO0332; LIM; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
DR LIM domain; Metal-Binding; Zinc.
SQ SEQUENCE 753 AA; 84059 MW; 79F98E47C100CF22 CRC64;

Query Match 9.1%; Score 69; DB 11; Length 753;
Best Local Similarity 100.0%; Pred. No. 1.7e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 CVECQKTVPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYICKPHFNQLFKSK 449
DB 388 CVECQKTVPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYICKPHFNQLFKSK 447
QY 450 GNYDEGFGH 458
DB 450 GNYDEGFGH 458
```

```
Db 448 GNYDEGFGH 456

RESULT 4
Q8C7S2 PRELIMINARY; PRT; 753 AA.
ID Q8C7S2;
AC Q8C7S2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK049350; BAC33699.1; -.
SQ SEQUENCE 753 AA; 84089 MW; 39571A17DF21F2C0 CRC64;

Query Match 9.1%; Score 69; DB 11; Length 753;
Best Local Similarity 100.0%; Pred. No. 1.7e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 CVECQKTVPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYICKPHFNQLFKSK 449
DB 388 CVECQKTVPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYICKPHFNQLFKSK 447
QY 450 GNYDEGFGH 458
DB 448 GNYDEGFGH 456

RESULT 5
Q8C3R7 PRELIMINARY; PRT; 593 AA.
ID Q8C3R7;
AC Q8C3R7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Epithelial protein lost in neoplasm.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK085065; BAC39353.1; -.
SQ SEQUENCE 593 AA; 66057 MW; FFB3E6B9506A61A CRC64;

Query Match 8.6%; Score 65; DB 11; Length 593;
Best Local Similarity 100.0%; Pred. No. 2.3e-59;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 CVECQKTVPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYICKPHFNQLFKSK 449
DB 228 CVECQKTVPMERLLANQOVFHISCFRCSCYNNKLSLGTYSALHGRYICKPHFNQLFKSK 287
QY 450 GNYDE 454
```

```
Db 288 GNYDE 292
|||||
RESULT 6
Q8N7Z0 PRELIMINARY; PRT; 519 AA.
AC Q8N7Z0
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ40200.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Wagaatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
RA Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.;
RA "NEDO human cDNA sequencing project.";
RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
DR EMBL: AK097519; BAC05086.1; -.
DR InterPro: IPR001781; LIM.
DR Pfam: PF00412; LIM; 1.
DR ProDom: PD000094; LIM; 1.
DR SMART: SM00132; LIM; 1.
DR PROSITE: PS00478; LIM_DOMAIN_1; 1.
DR PROSITE: PS00023; LIM_DOMAIN_2; 1.
KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 519 AA; 58797 MW; 69DE84BBE60E9106 CRC64;

Query Match 2.0%; Score 15; DB 4; Length 519;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFGH 458
|||||
Db 456 QLFKSKGNYDEGFGH 470

RESULT 7
Q8HXJ9 PRELIMINARY; PRT; 951 AA.
AC Q8HXJ9
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Medulla oblongata;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Sugano S.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Medulla oblongata;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,
RA Hirai M., Tetso K., Suzuki Y., Sugano S., Hashimoto K.;
RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
```

```
chromosomes.";
RL Gene 275:31-37(2001).
DR EMBL: AB097518; BAC41743.1; -.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 107682 MW; 42A3F8DEC05274E2 CRC64;

Query Match 2.0%; Score 15; DB 6; Length 951;
Best Local Similarity 100.0%; Pred. No. 3.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFGH 458
|||||
Db 469 QLFKSKGNYDEGFGH 483

RESULT 8
Q8BT23 PRELIMINARY; PRT; 127 AA.
AC Q8BT23
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ34982.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Wagaatsuma M.,
RA Wagaatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.;
RA "NEDO human cDNA sequencing project.";
RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
DR EMBL: BC004400; AAH04400.1; -.
DR EMBL: AK092301; BAC03855.1; -.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001781; LIM.
DR Pfam: PF00412; LIM; 1.
DR ProDom: PD000094; LIM; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS00478; LIM_DOMAIN_1; 1.
DR PROSITE: PS00023; LIM_DOMAIN_2; 1.
KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 127 AA; 14070 MW; B63174FCF0486956 CRC64;

Query Match 1.8%; Score 14; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 5.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFGH 457
|||||
Db 94 QLFKSKGNYDEGFGH 107

RESULT 9
Q96S91 PRELIMINARY; PRT; 127 AA.
AC Q96S91
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
```


DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Li N., Zhang W., Zhang M., Wan T., Cao X.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
DR EMBL; AY037154; AAK67634.1; --
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS00023; LIM_DOMAIN_2; 1.
DR KW Hypothetical protein; LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 127 AA; 14142 MW; B63014FEF0486954 CRC64;

Query Match 1.8%; Score 14; DB 4; Length 127;
Best Local Similarity 100.0%; Pred. No. 5.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFG 457
Db 94 QLFKSKGNYDEGFG 107
|||||

RESULT 10
Q8BGB5 PRELIMINARY; PRT; 128 AA.
ID Q8BGB5
AC Q8BGB5
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DT Hypothetical LIM domain/LIM domain profile/cytochrome c family
DE heme-binding site containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Body, Hippocampus, and Olfactory brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK012581; BAC25371.1; --
DR EMBL; AK032430; BAC27866.1; --
DR EMBL; AK049809; BAC33928.1; --
KW Hypothetical protein.
SQ SEQUENCE 128 AA; 14237 MW; ADF9161771331D13 CRC64;

Query Match 1.8%; Score 14; DB 11; Length 128;
Best Local Similarity 100.0%; Pred. No. 5.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 QLFKSKGNYDEGFG 457
Db 95 QLFKSKGNYDEGFG 108
|||||

RESULT 11
Q9DEY8

ID Q9DEY8 PRELIMINARY; PRT; 629 AA.
AC Q9DEY8
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DT Cytokeleton-associated LIM domain protein.
GN EPLIN.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
ON NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21100452; PubMed=11179579;
RA Maul R.S., Sachi Gerbin C., Chang D.D.;
RT "Characterization of mouse epithelial protein lost in neoplasm (EPLIN)
RT and comparison of mammalian and zebrafish EPLIN."
RL Gene 262:155-160(2001).
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
DR EMBL; AF307846; AAG31149.1; --
DR HSP; P32985; ICTU.
DR ZFIN; ZDB-GENE-001120-1; eplin.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS00023; LIM_DOMAIN_2; 1.
DR KW LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 629 AA; 70037 MW; C42341B024818C03 CRC64;

Query Match 1.8%; Score 14; DB 13; Length 629;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 KGYDEGFGHRPHK 462
Db 345 KGYDEGFGHRPHK 358
|||||

RESULT 12
Q9DWB9 PRELIMINARY; PRT; 647 AA.
ID Q9DWB9
AC Q9DWB9
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE PR83.
GN R83.
OS Rat cytomegalovirus (strain Maastricht).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Muromegalovirus.
ON NCBI_TaxID=79700;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Maastricht;
RX MEDLINE=20366325; PubMed=10906222;
RA Vink C., Beuken E., Bruggeman C.A.;
RT "Complete DNA sequence of the rat cytomegalovirus genome."
RL J. Virol. 74:7656-7665(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Maastricht;
RX MEDLINE=20473137; PubMed=11018281;
RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A., Vink C.;
RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a
RT spliced transcript."
RL Virus Res. 69:119-130(2000).
DR EMBL; AF232689; AAF99171.1; --
SQ SEQUENCE 647 AA; 71565 MW; 8C18506591522F6C CRC64;

Query Match 1.2%, Score 9, DB 12, Length 647;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 751 YYDEDEDEE 759
DB 447 YYDEDEDEE 455

RESULT 13

Q62493 PRELIMINARY; PRT; 101 AA.
AC Q62493
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Unknown protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cochlea;
RX MEDLINE=97237053; PubMed=9119401;
RA Crozet F., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA Hamel C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.G.,
RA Weil D., Pujol R., Petit C.;
RT "Cloning of the genes encoding two murine and human cochlear
RT unconventional type I myosins.";
RL Genomics 40:332-341(1997).
DR EMBL; 278149; CAB01551.1; -.
FT NON_TER 1 1
FT NON_TER 101 101
SQ SEQUENCE 101 AA; 11373 MW; 32A53AEC6AA7C477 CRC64;

Query Match 1.1%; Score 8; DB 11; Length 101;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 LSPDSRAS 368
DB 81 LSPDSRAS 88

RESULT 14

Q9ALE6 PRELIMINARY; PRT; 120 AA.
AC Q9ALE6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Dissimilatory sulfite reductase subunit A (fragment).
GN DSR.
OS uncultured sulfate-reducing bacterium UMTADSr828-28.
OC Bacteria; Proteobacteria; Deltaproteobacteria; environmental samples.
OX NCBI_TaxID=151117;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21318708; PubMed=11425735;
RA Chang Y.J., Peacock A.D., Long P.E., Stephen J.R., McKinley J.P.,
RA Macnaughton S.J., Hussain A.K., Saxton A.M., White D.C.;
RT "Diversity and characterization of sulfate-reducing bacteria in
RT groundwater at a uranium mill tailings site";
RL Appl. Environ. Microbiol. 67:3149-3160(2001).
DR EMBL; AY015533; AAG61216.1; -.
DR InterPro; IPR006067; Nlr_Sir_4Fe4S.
DR Pfam; PF01077; Nlr_Sir; 1.
FT NON_TER 1 1
FT NON_TER 120 120
SQ SEQUENCE 120 AA; 13189 MW; 328A679BDABFA964 CRC64;

Query Match 1.1%; Score 8; DB 2; Length 120;

Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 538 LGSAGSAL 545
DB 84 LGSAGSAL 91

RESULT 15

Q97214 PRELIMINARY; PRT; 122 AA.
AC Q97214;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein SS00922.
GN SS00922.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=1142726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Etrauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006712; AAK41204.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 122 AA; 14725 MW; 944101A92BD87F15 CRC64;

Query Match 1.1%; Score 8; DB 17; Length 122;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 IEKYNVPL 194
DB 37 IEKYNVPL 44

Search completed: January 6, 2004, 09:53:07
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2004, 09:49:46 ; Search time 13111 Seconds
(without alignments)
11560.534 Million cell updates/sec

Title: US-09-890-549-16
Perfect score: 3705
Sequence: 1 ggcgcagcagcagtaggtg.....atttgaaaaa..... 3705

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3660.4	98.8	3664	6	AX086386	AX086386 Sequence
2	3660.4	98.8	3664	9	HSN800164	AL136911 Homo sapi
3	3643.8	98.3	3655	9	AF198454	AF198454 Homo sapi
4	3374.4	91.1	3550	9	AF198455	AF198455 Homo sapi
5	3262.2	88.0	3277	9	AK000335	AK000335 Homo sapi
6	2698.2	72.8	2783	6	AX017487	AX017487 Sequence
7	2698.2	72.8	2783	6	BD135177	BD135177 Human nuc
8	2680	72.3	3348	9	AF157325	AF157325 Homo sapi
9	2604.4	70.3	2667	9	AK000057	AK000057 Homo sapi
10	2515.4	67.9	2536	9	AK000372	AK000372 Homo sapi
11	2284.4	61.7	115345	9	AC008147	AC008147 Homo sapi
12	2141	57.8	2164	9	BC001247	BC001247 Homo sapi
13	2115.2	57.1	2121	9	BC010664	BC010664 Homo sapi
14	1721.4	46.5	1754	9	AF218025	AF218025 Homo sapi
15	1568.4	42.3	2207	6	BD160117	BD160117 Primer fo
16	1568.4	42.3	2207	9	AK023649	AK023649 Homo sapi
17	1560.8	42.1	2793	9	AK096172	AK096172 Homo sapi
18	1559	42.1	3232	10	BC031490	BC031490 Mus muscu
19	1514.6	40.9	2370	10	AF307845	AF307845 Mus muscu
20	1409.6	38.0	3997	10	AF307844	AF307844 Mus muscu
21	1158.4	31.3	28186	9	AC139016	AC139016 Homo sapi
22	991	26.7	1021	6	AX012180	AX012180 Sequence
23	806.4	21.8	4425	10	CGU22818	G56637 SHGC-102015
24	766.4	20.7	777	11	G56637	U22818 Cricetus
25	669.6	18.1	195301	2	AC134548	AC134548 Mus muscu
26	637.4	17.2	641	6	AX011639	AX011639 Sequence
27	605.4	16.3	237661	2	AC114454	AC114454 Rattus no
28	416	11.2	532	6	BD154971	BD154971 Primer fo
29	390	10.5	390	6	BD113230	BD113230 EST and e
30	377.6	10.2	389	6	BD025824	BD025824 Sequence
31	352.6	9.5	413	6	AR270117	AR270117 Sequence
32	343	9.3	343	6	AX378979	AX378979 Sequence
33	343	9.3	343	6	AX379294	AX379294 Sequence
34	310.2	8.4	195516	2	AC135476	AC135476 Rattus no
35	277.4	7.5	279	6	AR269930	AR269930 Sequence
36	240.8	6.5	250	11	G11512	G11512 human STS S
37	224	6.0	238	6	AR269818	AR269818 Sequence
38	212	5.7	247	6	AX184947	AX184947 Sequence
39	210.8	5.7	231	6	AX187815	AX187815 Sequence
40	210.8	5.7	255	6	AX186750	AX186750 Sequence
41	209.4	5.7	265	6	AX098162	AX098162 Sequence
42	199.2	5.4	296	6	BD027230	BD027230 Sequence
43	175.2	4.7	178	6	AX341839	AX341839 Sequence
44	147.4	4.0	2264	9	AK097519	AK097519 Homo sapi
45	145.8	3.9	3287	9	AB097518	AB097518 Macaca fa

ALIGNMENTS

RESULT 1
AX086386
LOCUS AX086386 3664 bp DNA linear PAT 09-MAR-2001
DEFINITION Sequence 338 from Patent WO0112659.
ACCESSION AX086386
VERSION AX086386.1 GI:13275951
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Wiemann,S.
TITLE Human dna sequences
JOURNAL Patent: WO 0112659-A 338 22-FEB-2001;

Pred. No. is the number of results predicted by chance to have a

German Human Genome Project (DE)	
Location/Qualifiers	
1..3664	
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
BASE COUNT	1210 a 734 c 809 g 911 t
ORIGIN	
Query Match	98.8%; Score 3660.4; DB 6; Length 3664;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 3661; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	44 GGCTAGGTAGACGCCGGACCTGTGACAGGCTGTGAGCGCTGTGAGAGAAAGCGGC 103
Db	1 GGCTAGGTAGACGCCGGACCTGTGACAGGCTGTGAGCGCTGTGAGAGAAAGCGGC 60
QY	104 TTTTASCCAGGTATTTTCAGTCTCTGTAGACAAGATGGAATCATCTCCATTTATACAGG 163
Db	61 TTTTACCCAGGTATTTTCAGTCTCTGTAGACAAGATGGAATCATCTCCATTTAATAGCGG 120
QY	164 CAATGGACCTCACTATCATTTAGGGTAAACAGCCAAAGAACTTTCTCTGTCAACAAGAAC 223
Db	121 CAATGGACCTCACTATCATTTAGGGTAAACAGCCAAAGAACTTTCTCTGTCAACAAGAAC 180
QY	224 AAGTCATCGGCTATTTGTGAAATATTTCTCAAGTACCAAGAACAGCTGAAGAAAACAAAC 283
Db	181 AAGTCATCGGCTATTTGTGAAATATTTCTCAAGTACCAAGAACAGCTGAAGAAAACAAAC 240
QY	284 ATGGAGAAAGAGAGTAATAACCCGAAATCTCTCCAGCACTTTAGAAAGGGGACCCCTG 343
Db	241 ATGGAGAAAGAGAGTAATAACCCGAAATCTCTCCAGCACTTTAGAAAGGGGACCCCTG 300
QY	344 ACTGTGTTAAAGAAAGTGGGAGAACCCAGGCTGGGAGCAGAGTCTCACACAGACTCT 403
Db	301 ACTGTGTTAAAGAAAGTGGGAGAACCCAGGCTGGGAGCAGAGTCTCACACAGACTCT 360
QY	404 CTACGGAACAGCAGCAGCTGAGATTAGGACACAGCAGACCAATCTCTCTGAGTGACA 463
Db	361 CTACGGAACAGCAGCAGCTGAGATTAGGACACAGCAGACCAATCTCTCTGAGTGACA 420
QY	464 AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGA 523
Db	421 AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGA 480
QY	524 CTCAGTCACTCTCTGAAGCCCTCTGTTTCAGGTCGATATCCCCACATCAAGGACGGTGAG 583
Db	481 CTCAGTCACTCTCTGAAGCCCTCTGTTTCAGGTCGATATCCCCACATCAAGGACGGTGAG 540
QY	584 GATCTTAAAGACCTCAACAGAAAGTAAAAAATGGAAAAATTGTCTAGGAGAAATCCAGG 643
Db	541 GATCTTAAAGACCTCAACAGAAAGTAAAAAATGGAAAAATTGTCTAGGAGAAATCCAGG 600
QY	644 CATGAAGTAGAAAAATCAGAAATCAGTGAATAACACAGATGCTTTCGGGCAAAATAGAGAAA 703
Db	601 CATGAAGTAGAAAAATCAGAAATCAGTGAATAACACAGATGCTTTCGGGCAAAATAGAGAAA 660
QY	704 TATAATGTTCCGCTGAAACAGGCTTAAAGATGATGTTTGAGAAAGGTGAACCAACTCAAACT 763
Db	661 TATAATGTTCCGCTGAAACAGGCTTAAAGATGATGTTTGAGAAAGGTGAACCAACTCAAACT 720
QY	764 AAGATTCTCCGGGCCAAAGCCGAAGTGAAGTGGAAAGGAGATCTCTGAAAAACAGCTAT 823
Db	721 AAGATTCTCCGGGCCAAAGCCGAAGTGAAGTGGAAAGGAGATCTCTGAAAAACAGCTAT 780
QY	824 TCTCTAGATGACCTGGAAATAGGCCAGGTCAGTTGTCATCTTTCATCTTTCAGCTCGGAG 883
Db	781 TCTCTAGATGACCTGGAAATAGGCCAGGTCAGTTGTCATCTTTCATCTTTCAGCTCGGAG 840
QY	884 AAAATAGAGTAGACGAAATCTGAACTTCCACGCTCTCAGAAACCTCTATAAAGAT 943
Db	841 AAAATAGAGTAGACGAAATCTGAACTTCCACGCTCTCAGAAACCTCTATAAAGAT 900

QY	944	CGAATGCCAAGTACACAGGCGAGTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAAT	1003
Db	901	CGAATGCCAAGTACACAGGCGAGTGTGTCCAAACAAAGCAGCTCAACCAACTATACAAAT	960
QY	1004	GAGCTGAAGCCAGTGTGGGAAATCAAAATTCATAAAATGAGCAAAAGGAGATGTG	1063
Db	961	GAGCTGAAGCCAGTGTGGGAAATCAAAATTCATAAAATGAGCAAAAGGAGATGTG	1020
QY	1064	CCCCAGGTCTGTAGGTCTGCATCACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAG	1123
Db	1021	CCCCAGGTCTGTAGGTCTGCATCACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAG	1080
QY	1124	AATAGCCTGCGAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTGACTCCCAAGTTAAG	1183
Db	1081	AATAGCCTGCGAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTGACTCCCAAGTTAAG	1140
QY	1184	AGTGAGGTTCAACAGCCTGTCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCCAGT	1243
Db	1141	AGTGAGGTTCAACAGCCTGTCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCCAGT	1200
QY	1244	CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACC	1303
Db	1201	CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACC	1260
QY	1304	TGCGTGGAAATGTCAGAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAGGTG	1363
Db	1261	TGCGTGGAAATGTCAGAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAGGTG	1320
QY	1364	TTTCACATCAGCTGCTTCCGTTGCTCTCTATTCGCAACAAACCTCAGTCTAGGAACATAT	1423
Db	1321	TTTCACATCAGCTGCTTCCGTTGCTCTCTATTCGCAACAAACCTCAGTCTAGGAACATAT	1380
QY	1424	GCATCTTTACATCGAAGAACTATTGTAAGCCTCAGTTCAATCAACTCTTTAAATCTAAG	1483
Db	1381	GCATCTTTACATCGAAGAACTATTGTAAGCCTCAGTTCAATCAACTCTTTAAATCTAAG	1440
QY	1484	GGCAACTATGATGAAGGCTTTGGGCACAGACACAAAGATCTATGGCAAGCAAAAT	1543
Db	1441	GGCAACTATGATGAAGGCTTTGGGCACAGACACAAAGATCTATGGCAAGCAAAAT	1500
QY	1544	GAAAAAGAAAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGAGAGACCCCTCAC	1603
Db	1501	GAAAAAGAAAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGAGAGACCCCTCAC	1560
QY	1604	AGCCAGGGGTAGAAATGCCCCATTGCTAAGTGGGTGTCTCTGGCTCAAGTATGGAA	1663
Db	1561	AGCCAGGGGTAGAAATGCCCCATTGCTAAGTGGGTGTCTCTGGCTCAAGTATGGAA	1620
QY	1664	GCCAAAGCCCTCTCTCAGCAGGAGAGGAGACAAAGCAGCTGAAACCAAGAGCTGAGG	1723
Db	1621	GCCAAAGCCCTCTCTCAGCAGGAGAGGAGACAAAGCAGCTGAAACCAAGAGCTGAGG	1680
QY	1724	ATCGCTGGCCACCCCTCACTGAACTTGGAAAGTTGAGGAAGTCCCTTGGAGAAAGGATC	1783
Db	1681	ATCGCTGGCCACCCCTCACTGAACTTGGAAAGTTGAGGAAGTCCCTTGGAGAAAGGATC	1740
QY	1784	AAATGTCAAAAGCCCAAAATGGGCTCTGTAAGACGAAATCAGCAAGCCCGAAAGTCTCTGAG	1843
Db	1741	AAATGTCAAAAGCCCAAAATGGGCTCTGTAAGACGAAATCAGCAAGCCCGAAAGTCTCTGAG	1800
QY	1844	GATGTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGAAAGAGCCGC	1903
Db	1801	GATGTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGAAAGAGCCGC	1860
QY	1904	CCATTCACTGTAGCAGCTTCAATTTCAAAGCACCTCTGTCAGAGGCCCAAAACTCTGTCC	1963
Db	1861	CCATTCACTGTAGCAGCTTCAATTTCAAAGCACCTCTGTCAGAGGCCCAAAACTCTGTCC	1920
QY	1964	CCACTTATCAGAAAGGCTTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAAGA	2023
Db	1921	CCACTTATCAGAAAGGCTTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAAGA	1980
QY	2024	GTTCAGAAAGGAAACAAAGTGGAAAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTGGGA	2083

Db 1981 GTTGCAAGAGGAAACAAGTGGAAATGCAAGGCTTCTAAGAGAAATGGGAATGTGGGA 2040
Qy 2084 ABAACACCTGGCAAAACAAGAAATCTAAGAGAGACAGAGGGAAGAGTAAGTAAGGT 2143
Db 2041 AAAACAACCTGGCAAAACAAGAAATCTAAGAGAGACAGAGGGAAGAGTAAGTAAGGT 2100
Qy 2144 CATAGTTTGGAGATGGAGATGAGAAATCTTGTAGAAAATGGTGCACACTCCGATGAAGAT 2203
Db 2101 CATAGTTTGGAGATGGAGATGAGAAATCTTGTAGAAAATGGTGCACACTCCGATGAAGAT 2160
Qy 2204 GATAACAGCTTCTCAAAACAACAAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCGAGT 2263
Db 2161 GATAACAGCTTCTCAAAACAACAAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCGAGT 2220
Qy 2264 TTTGTAGACAACACCTTTTGGTGAAGAAATTCACCTACTCAGAAATCCAGGATGTG 2323
Db 2221 TTTGTAGACAACACCTTTTGGTGAAGAAATTCACCTACTCAGAAATCCAGGATGTG 2280
Qy 2324 GAACTCTGGAGGAGAGAGTGGTCAAGAGCTCTGTGGAGAAACAGATAAAGAGAAAT 2383
Db 2281 GAACTCTGGAGGAGAGAGTGGTCAAGAGCTCTGTGGAGAAACAGATAAAGAGAAAT 2340
Qy 2384 CGGTATTATGATGAGGATGAGGATGAAGAGTGAACAAATTCGAATGATGCTGGCCCTTAAA 2443
Db 2341 CGGTATTATGATGAGGATGAGGATGAAGAGTGAACAAATTCGAATGATGCTGGCCCTTAAA 2400
Qy 2444 TTCATGTTAGTGTAGCGACCACTGCCCTTTGTCAAAAATGTGATGCACATTAAGCAGGTA 2503
Db 2401 TTCATGTTAGTGTAGCGACCACTGCCCTTTGTCAAAAATGTGATGCACATTAAGCAGGTA 2460
Qy 2504 TCCAGCATGAATGTAAATTTACTTGGAGTAACCTTTGGAAGAAATTCCTTCTTAAAT 2563
Db 2461 TCCAGCATGAATGTAAATTTACTTGGAGTAACCTTTGGAAGAAATTCCTTCTTAAAT 2520
Qy 2564 CAAAAACAAAAACAAAAACAAAAACACATTTCTAAATCTAGAGATAAATTTTACTTTA 2623
Db 2521 CAAAAACAAAAACAAAAACAAAAACACATTTCTAAATCTAGAGATAAATTTTACTTTA 2580
Qy 2624 AATTCCTCAATTTTACGAGTATGATATGATGCTGAAGCTTTGAAGCTGGGGA 2683
Db 2581 AATTCCTCAATTTTACGAGTATGATATGATGCTGAAGCTTTGAAGCTGGGGA 2640
Qy 2684 TATTCACCTGATAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAAGTGA 2743
Db 2641 TATTCACCTGATAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAAGTGA 2700
Qy 2744 TAGATGATTAGTAGTATTTGTACACACTATTTTGGAAATAGAGACATACAGAGGAA 2803
Db 2701 TAGATGATTAGTAGTATTTGTACACACTATTTTGGAAATAGAGACATACAGAGGAA 2760
Qy 2804 TTTAGGGGCTTAAACATTACGACTGAATGCATTTAGTATAAAGGCGACAGTTTGTATAT 2863
Db 2761 TTTAGGGGCTTAAACATTACGACTGAATGCATTTAGTATAAAGGCGACAGTTTGTATAT 2820
Qy 2864 TTTTAAATGAATACCAATTTAATTTTGTAGTATTTTACCTGTTTAAAGAGATTTTGTCTT 2923
Db 2821 TTTTAAATGAATACCAATTTAATTTTGTAGTATTTTACCTGTTTAAAGAGATTTTGTCTT 2880
Qy 2924 TAAATTTTTAGGTTAATTTTCTGCTGTATATATATAGAGAAATTTACTATTTATGTC 2983
Db 2881 TAAATTTTTAGGTTAATTTTCTGCTGTATATATATAGAGAAATTTACTATTTATGTC 2940
Qy 2984 CTGCTCTCTAAACTACATCCTGAATCGACGCTCGAGTATAATACACAGAGACATTT 3043
Db 2941 CTGCTCTCTAAACTACATCCTGAATCGACGCTCGAGTATAATACACAGAGACATTT 3000
Qy 3044 TTGAGGCAATTTGAAAAACCAACTACCTCTTTCGGTGTCTTAGAGAGATCTGCTGCTCCC 3103
Db 3001 TTGAGGCAATTTGAAAAACCAACTACCTCTTTCGGTGTCTTAGAGAGATCTGCTGCTCCC 3060
Qy 3104 AAATAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTCCTTTCTTCT 3163

3061 AAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTGCTTTCTTCT 3120
Qy 3164 GGTGATATCTGTGCTTCTCATAAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGG 3223
Db 3121 GGTGATATCTGTGCTTCTCATAAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGG 3180
Qy 3224 ATCACTGTCCCCCTCTTCCGCTGTAGAGCAAGTGAAGAGCTTTAAAGAGGAGAGAA 3283
Db 3181 ATCACTGTCCCCCTCTTCCGCTGTAGAGCAAGTGAAGAGCTTTAAAGAGGAGAGAA 3240
Qy 3284 AGAACTGTCTTACACCACTTTGAGCTCAGACCTCTAAACCCCTGTATTTCCCTTATGATGC 3343
Db 3241 AGAACTGTCTTACACCACTTTGAGCTCAGACCTCTAAACCCCTGTATTTCCCTTATGATGC 3300
Qy 3344 CCCTTTTGTAGACACTAAATTTTAAATCTTACTAGCTCTGAAATATATTTGATTTTATC 3403
Db 3301 CCCTTTTGTAGACACTAAATTTTAAATCTTACTAGCTCTGAAATATATTTGATTTTATC 3360
Qy 3404 ACAGTATTTCTCAGGCTGAATTTAAACCMACCTATAGGCCCTTTTCTTGGGATGATTTCTA 3463
Db 3361 ACAGTATTTCTCAGGCTGAATTTAAACCMACCTATAGGCCCTTTTCTTGGGATGATTTCTA 3420
Qy 3464 GTCTTAAGGTTTGGGACATTTATAAATCTTGAAGCTTTGAGTACATTTTGTGTACACAGTTGATTTCC 3523
Db 3421 GTCTTAAGGTTTGGGACATTTATAAATCTTGAAGCTTTGAGTACATTTTGTGTACACAGTTGATTTCC 3480
Qy 3524 AAATTTGATGATGGAGGGAGAGGTGTCTTTAAAGCTGTAGGCTTTTCTTGTACTGCAAT 3583
Db 3481 AAATTTGATGATGGAGGGAGAGGTGTCTTTAAAGCTGTAGGCTTTTCTTGTACTGCAAT 3540
Qy 3584 TATAGAGATTTAGCTTTAATAATTTTTTAGAGATGTAAACATCTCTGCTTTCTTGTACTT 3643
Db 3541 TATAGAGATTTAGCTTTAATAATTTTTTAGAGATGTAAACATCTCTGCTTTCTTGTACTT 3600
Qy 3644 CCTAGCTGAACATTTTATTTCAATAAGATTTTAAATTTGAAAAAATTTGAAAAAATTTT 3703
Db 3601 CCTAGCTGAACATTTTATTTCAATAAGATTTTAAATTTGAAAAAATTTTAAATTTT 3660
Qy 3704 AA 3705
Db 3661 AA 3662

RESULT 2
HSM800164
LOCUS HSM800164 3664 bp mRNA linear PRI 03-AUG-2001
DEFINITION Homo sapiens mRNA; cDNA DKFZp586i1918 (from clone DKFZp586i1918);
complete cds.
ACCESSION AL136911
VERSION AL136911.1 GI:12053316
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3664)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2001) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp586i1918) is available at the R2PD in Berlin.
Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clonesrzd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.
Location/Qualifiers
source 1..3664
/organism="Homo sapiens"
/mol_type="mRNA"

/db_xref="taxon:9606"
/map="12"
/clone="DKFZp58611918"
/issue_type="uterus"
/clone_lib="586 (synonym: hute1). Vector pSport1; host
DH10B; sites NotI + SalI/MluI"
/dev_stage="adult"
1. 3664
/gene="DKFZp58611918"
94...2373
/gene="DKFZp58611918"
/note="similarity to transcription factor SF3"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB56845.1"
/db_xref="GI:12053317"
/db_xref="SPTREMBL:Q5UH66"
/translation="MESSPNNRQWTSLSRVIAKELSLVKNKSSAIVEIFSKYOKA
AEETNMBKRGNTENLSOHEFKGLTLVKKKNWPGIAGSHDLSLNSSTETRHAD
HPAEVTSHAASGAKAQDEOIHPRSLRSPPEALVOGRYPHIKDGEDLDKHTEKK
MENCLGSRHEVEKSEISENTDASGKIEKYNPLNLRKMFKEGPTQTKILRAQSR
ASVKRISNSYSLDLIGPQLSSSTFDSEKNESRRNLELPLRSETSIKDRMAKQA
AVKQSSSTNTYSLKASGGEIKHKMEQKENVPPPEVCI THQGEKIKI SANNESLAV
RSTPAEDSRDSQVSKVQVPHKPLSPDRASSLSSESPKAMKKFOAPARETCVE
CKTQVYMERLLANQOVPHISCFRCSCYNNKLSIGTVASLHGRIYCKPHENQLPKSG
NYDEGFGHRPHKDIAWASKNENEILERPQALANARETPHSPGVEDAPIAKVGLAASM
EAKASSQEKEDKPAETKRIAWPPPTTELGSSALSEEGIKMSKPKWPEDEISKPE
VPEDVDLKLRLSSLSKERSRPTVAASFQSTSVKSPKTVSPPIRKWMSGEQSEE
SVGRVAERKQVENAKSKNGNVKTTWQNKESGETKPKSKEGSHLENNENLVEN
GADSEDDNFLKQSQPOEKPSLWSSFVNDTFAEEFTTQNKSQDVLEWEGEVKEL
SVEEQIKRNYRDEDEEE"
3637, 3642
polyA_signal
/gene="DKFZp58611918"
3647
polyA_site
/gene="DKFZp58611918"
BASE COUNT 1210 a 734 c 809 g 911 t
ORIGIN
Query Match 98.8%; Score 3660.4; DB 9; Length 3664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3661; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 44 GCGCTAGGTAGAGCGCGGACCTGTGACAGGGGTGTGACAGGCGAGAGAAAGCGGC 103
DB 1 GCGGTAGGTAGAGCGCGGACCTGTGACAGGGGTGTGACAGGCGAGAGAAAGCGGC 60
QY 104 TTTTAGCCAGGTATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCCATTATATACGG 163
DB 61 TTTTAGCCAGGTATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCCATTATATACGG 120
QY 164 CAATGGACCTCACTATCATTTGAGGGTACAGCCAAAGAACTTTCTCTTGTCAACAAGAAC 223
DB 121 CAATGGACCTCACTATCATTTGAGGGTAAACAGCCAAAGAACTTTCTCTTGTCAACAAGAAC 180
QY 224 AAGTCATCGGCTATTGGAATATTTCTCCAAGTACCAGAAAGCAGCTGAAGAAACAAC 283
DB 181 AAGTCATCGGCTATTGGAATATTTCTCCAAGTACCAGAAAGCAGCTGAAGAAACAAC 240
QY 284 ATGGAGAAGAGAGAGTAAACCGGAAATCTCTCCAGACATTTAGAAAGGGGACCCCTG 343
DB 241 ATGGAGAAGAGAGAGTAAACCGGAAATCTCTCCAGACATTTAGAAAGGGGACCCCTG 300
QY 344 ACTGTGTTAAGAGAGTGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCT 403
DB 301 ACTGTGTTAAGAGAGTGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCT 360
QY 404 CTACGGACAGCAGCAGTATGAGTATAGGCACAGAGCAGACATCTCTCTGCTGAAAGTGACA 463
DB 361 CTACGGACAGCAGCAGTATGAGTATAGGCACAGAGCAGACATCTCTCTGCTGAAAGTGACA 420
QY 464 AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGACAAATCCACCCAGACTCTAGA 523
DB 421 AGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAGACAAATCCACCCAGACTCTAGA 480

QY 524 CT CAGGTACCTCTCTGAAGCCCTGTTTCAGGGTCGATATCCCCACATCAAGGACGGTGAG 583
DB 481 CT CAGGTACCTCTCTGAAGCCCTGTTTCAGGGTCGATATCCCCACATCAAGGACGGTGAG 540
QY 584 GATCTTTAAAGACCACTCAACAGAAAGTAAAAAATGGAAAAATGTTCTAGAGAAATCCAGG 643
DB 541 GATCTTTAAAGACCACTCAACAGAAAGTAAAAAATGGAAAAATGTTCTAGAGAAATCCAGG 600
QY 644 CATGAAGTAGAAAAATCAGAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAA 703
DB 601 CATGAAGTAGAAAAATCAGAAATCAGTGAACACACAGATGCTTCGGGCAAAATAGAGAAA 660
QY 704 TATAATGTTCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGGTGAACCAACTCAAACT 763
DB 661 TATAATGTTCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGGTGAACCAACTCAAACT 720
QY 764 AAGATTCTCGGGCCCCAAAGCGGAAGTGAAGTGAAGAAAGATCTCTGAAAAACAGCTAT 823
DB 721 AAGATTCTCGGGCCCCAAAGCGGAAGTGAAGTGAAGAAAGATCTCTGAAAAACAGCTAT 780
QY 824 TCTCTAGATGACTGTGAATAGGCCAGGTGAGTCTGATCTTCTACATTTTCACTTCGAGT 883
DB 781 TCTCTAGATGACTGTGAATAGGCCAGGTGAGTCTGATCTTCTACATTTTCACTTCGAGT 840
QY 884 AAAAAATGAGTAGACGAAATCTGGAATCTCCACGCTCTCAGAAAACCTCTATAAAGAT 943
DB 841 AAAAAATGAGTAGACGAAATCTGGAATCTCCACGCTCTCAGAAAACCTCTATAAAGAT 900
QY 944 CGAATGGCCAAAGTACAGGCGAGCTGTGTCAAAACAAAAGCAGCTCAACCAACTATACAA 1003
DB 901 CGAATGGCCAAAGTACAGGCGAGCTGTGTCAAAACAAAAGCAGCTCAACCAACTATACAA 960
QY 1004 GAGCTGAAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGAGAAATGTG 1063
DB 961 GAGCTGAAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGAGAAATGTG 1020
QY 1064 CCCCAGGTCTCTGAGTCTGCAATCAACCATCAGGAAGGGGAAAGATTTCTGCAATAG 1123
DB 1021 CCCCAGGTCTCTGAGTCTGCAATCAACCATCAGGAAGGGGAAAGATTTCTGCAATAG 1080
QY 1124 AATAGCCTGGCAGTCCGTTCCACCCCTCCGAAAGATGACTCCCGTGACTCCCGAGTTAAG 1183
DB 1081 AATAGCCTGGCAGTCCGTTCCACCCCTCCGAAAGATGACTCCCGTGACTCCCGAGTTAAG 1140
QY 1184 AGTAGGTTCAACAGCTGTCCATCCCAAGCACTAAGTCCAGATTCCAGAGCTCCAGT 1243
DB 1141 AGTAGGTTCAACAGCTGTCCATCCCAAGCACTAAGTCCAGATTCCAGAGCTCCAGT 1200
QY 1244 CTCTCTGAAAGTTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAGACC 1303
DB 1201 CTCTCTGAAAGTTCTCTCCCAAGCAATGAAGAGTTTCAGGCACCTGCAAGAGAGACC 1260
QY 1304 TCGTGGAAATGTGAGAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAGGTG 1363
DB 1261 TCGTGGAAATGTGAGAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAGGTG 1320
QY 1364 TTTCACTACAGTCTGTTCCGTTGCTCTATTGCAACCAACAACTCAGTCTAGGAAACATAT 1423
DB 1321 TTTCACTACAGTCTGTTCCGTTGCTCTATTGCAACCAACAACTCAGTCTAGGAAACATAT 1380
QY 1424 GCATCTTTACATGGAAGATCTATTGTAAGCCTCACTTCAATCAACTCTTTTAAATCTTAAG 1483
DB 1381 GCATCTTTACATGGAAGATCTATTGTAAGCCTCACTTCAATCAACTCTTTTAAATCTTAAG 1440
QY 1484 GGCAACTATGATGAAGGCTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAAAAAT 1543
DB 1441 GGCAACTATGATGAAGGCTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAAAAAT 1500
QY 1544 GAAAAACGAAGATTTTGGAGAGACCCAGCCAGCTTGCAAAATGCAAGGGAGAGCCCTCAC 1603
DB 1501 GAAAAACGAAGATTTTGGAGAGACCCAGCCAGCTTGCAAAATGCAAGGGAGAGCCCTCAC 1560

QY	1604	AGCCAGGGGTAGAA	GATGCCCTATTGCTAAGTGGGTCTCTGCTGCAAGTATGGAA	1663
Db	1561	AGCCAGGGGTAGAA	GATGCCCTATTGCTAAGTGGGTCTCTGCTGCAAGTATGGAA	1620
QY	1664	GCCAAAGCCTCTCT	CAGCAGGAGAGAGACAAGCCAGCTTAAACCAAGAAAGCTGAGG	1723
Db	1621	GCCAAAGCCTCTCT	CAGCAGGAGAGAGAGACAAGCCAGCTTAAACCAAGAAAGCTGAGG	1680
QY	1724	ATCGCTGGCCACCC	CCACTGAACTTGAAGTTTCAGGAAGTCCCTTGAGGAAGGGATC	1783
Db	1681	ATCGCTGGCCACCC	CCACTGAACTTGAAGTTTCAGGAAGTCCCTTGAGGAAGGGATC	1740
QY	1784	AAATGTCAAAGCC	CAATGGCCTCTGTAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAG	1843
Db	1741	AAATGTCAAAGCC	CAATGGCCTCTGTAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAG	1800
QY	1844	GATGTGATCTAG	ATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAGAACCCGC	1903
Db	1801	GATGTGATCTAG	ATCTGAAGAGCTTAAGACGATCTTCTTCACTGAAGGAAGAACCCGC	1860
QY	1904	CCATTCACTGTAC	GAGCTTCATTTCAAAGCACCTCTGTCAAAGGCCCAAAAACCTGTGTC	1963
Db	1861	CCATTCACTGTAC	GAGCTTCATTTCAAAGCACCTCTGTCAAAGGCCCAAAAACCTGTGTC	1920
QY	1964	CCACCTATCAG	GAAGGCTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGTGGAGA	2023
Db	1921	CCACCTATCAG	GAAGGCTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGTGGAGA	1980
QY	2024	GTTGCAGAAAG	GAACAACTGGAAGTTCCTAAGAAATGGGAATGGGA	2083
Db	1981	GTTGCAGAAAG	GAACAACTGGAAGTTCCTAAGAAATGGGAATGGGA	2040
QY	2084	AAAAACAACCT	GCCAAACAAAGAACTTAAGAGAGACAGAGGAAGAGAAAGT	2143
Db	2041	AAAAACAACCT	GCCAAACAAAGAACTTAAGAGAGACAGAGGAAGAGAAAGT	2100
QY	2144	CATAGTTTGG	AGATGAGAAATGAGATCTTTAGAAAATGGTGACACTCCGATCAAGAT	2203
Db	2101	CATAGTTTGG	AGATGAGAAATGAGATCTTTAGAAAATGGTGACACTCCGATCAAGAT	2160
QY	2204	GATAACAGCT	TCTCCTCAACAACTCTCCACAAAGAACCCCAAGTCTCTGAATTTGGTCGAGT	2263
Db	2161	GATAACAGCT	TCTCCTCAACAACTCTCCACAAAGAACCCCAAGTCTCTGAATTTGGTCGAGT	2220
QY	2264	TTTGTAGACAA	CACCTTTGCTGAAGAAATTCACACTCAGAAATCCAGAGATGTG	2323
Db	2221	TTTGTAGACAA	CACCTTTGCTGAAGAAATTCACACTCAGAAATCCAGAGATGTG	2280
QY	2324	GAACTCTGG	AGGAGAACTGTCTCAAGAGCTCTCTGTGGAGAACACAGATTAAGAGAAAT	2383
Db	2281	GAACTCTGG	AGGAGAACTGTCTCAAGAGCTCTCTGTGGAGAACACAGATTAAGAGAAAT	2340
QY	2384	CGGTATTAT	GATGAGGATGAGGATGAAGAGTGACAAATTTGCAATGATCTGGGCTTAAA	2443
Db	2341	CGGTATTAT	GATGAGGATGAGGATGAAGAGTGACAAATTTGCAATGATCTGGGCTTAAA	2400
QY	2444	TTTCATGTT	AGTGTAGAGCCACTGCCCTTTGTCAAAAATGTGATGCACATAAGCAGGTA	2503
Db	2401	TTTCATGTT	AGTGTAGAGCCACTGCCCTTTGTCAAAAATGTGATGCACATAAGCAGGTA	2460
QY	2504	TCCCAGCAT	GAAATGTAATTTACTTGAAGTAACCTTTGGAAAAAGAAATTCCTTCTTAAAT	2563
Db	2461	TCCCAGCAT	GAAATGTAATTTACTTGAAGTAACCTTTGGAAAAAGAAATTCCTTCTTAAAT	2520
QY	2564	CAAAAA	CAAAAAACACAAACACATTTCTAAATCTAGAGATACCTTTACTTTA	2623
Db	2521	CAAAAA	CAAAAAACACAAACACATTTCTAAATCTAGAGATACCTTTACTTTA	2580
QY	2624	AATTCCT	CAATTTAGCAGTGATATGCAATAGTGTGTAAAGCTTGTAACTGGGAAA	2683
Db	2581	AATTCCT	CAATTTAGCAGTGATATGCGTAAGTGTGTAAAGCTTGTAACTGGGAAA	2640
QY	2684	TATCCACCT	GATATAGCCAGATCTTACTGTATTTCCAAAAAGGCAATATTAAAGGTAGA	2743

RESULT 3

Db	2641	TATTCACCT	GATATAGCCAGATCTTACTGTATTTCCAAAAAGGCAATATTAAAGGTAGA	2700
QY	2744	TAGATGAT	TAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAAGGAA	2803
Db	2701	TAGATGAT	TAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAAGGAA	2760
QY	2804	TTTAGGGCT	TTAAACATTACGACTGAATGCACTTTAGTATAAAGGGCACAGTTTGATAT	2863
Db	2761	TTTAGGGCT	TTAAACATTACGACTGAATGCACTTTAGTATAAAGGGCACAGTTTGATAT	2820
QY	2864	TTTTAAAT	TGAATACCAATTTAAATTTTTTAGTATTTACCTGTTAAAGAGATTATTTAGTCTT	2923
Db	2821	TTTTAAAT	TGAATACCAATTTAAATTTTTTAGTATTTACCTGTTAAAGAGATTATTTAGTCTT	2880
QY	2924	TAAATTTTT	TAGGTTAATTTTTCTTGTGTATATATATGAGGAATTTACTACTTTATGTC	2983
Db	2881	TAAATTTTT	TAGGTTAATTTTTCTTGTGTATATATATGAGGAATTTACTACTTTATGTC	2940
QY	2984	CTGCTCTCT	TAAACTACATCTCGAACTCGACCTCTGAGGTATAATACAAAGAGACACTTT	3043
Db	2941	CTGCTCTCT	TAAACTACATCTCGAACTCGACCTCTGAGGTATAATACAAAGAGACACTTT	3000
QY	3044	TTGAGGCA	TAATGAAACCAACCTACACTCTTCGGTCTTTAGAGAGATCTGCTGTCTCC	3103
Db	3001	TTGAGGCA	TAATGAAACCAACCTACACTCTTCGGTCTTTAGAGAGATCTGCTGTCTCC	3060
QY	3104	AAATAAGCT	TTTTGTATCTGCCAGTGAATTTACTCTGACTCCAAATGATGCTTCTTTCT	3163
Db	3061	AAATAAGCT	TTTTGTATCTGCCAGTGAATTTACTCTGACTCCAAATGATGCTTCTTTCT	3120
QY	3164	GGTGATAT	CTGTCTCTCATTAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGG	3223
Db	3121	GGTGATAT	CTGTCTCTCATTAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGG	3180
QY	3224	ATCACTGT	CCCCCATCTTCGGTGTAGAGCAAAAGTGAAGAGTTTAAAGGAGGAAGAA	3283
Db	3181	ATCACTGT	CCCCCATCTTCGGTGTAGAGCAAAAGTGAAGAGTTTAAAGGAGGAAGAA	3240
QY	3284	AGAACTGT	CTTACACCACTTGAGCTCAGACCTCTAAACCTGTATTTCCCTTATGATGTC	3343
Db	3241	AGAACTGT	CTTACACCACTTGAGCTCAGACCTCTAAACCTGTATTTCCCTTATGATGTC	3300
QY	3344	CCCTTTT	TGAGACACTAAATTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT	3403
Db	3301	CCCTTTT	TGAGACACTAAATTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT	3360
QY	3404	ACAGTAT	TTCTCAGGGTGAATTTAAACCACTATAGGCCCTTTTCTTGGGATGATTTCTA	3463
Db	3361	ACAGTAT	TTCTCAGGGTGAATTTAAACCACTATAGGCCCTTTTCTTGGGATGATTTCTA	3420
QY	3464	GTCTTAAG	TTTGGGACATTTAAACTTGAGTACATTTGTTGTACACAGTTGATATTC	3523
Db	3421	GTCTTAAG	TTTGGGACATTTAAACTTGAGTACATTTGTTGTACACAGTTGATATTC	3480
QY	3524	AAATTTG	TATGATGGAGGGGAGAGGTGCTTAAGCTGTAGGCTTTCTTTGTACTGCAIT	3583
Db	3481	AAATTTG	TATGATGGAGGGGAGAGGTGCTTAAGCTGTAGGCTTTCTTTGTACTGCAIT	3540
QY	3584	TATAGAG	TTTACCTTTAATTTTTTAGAGATGTAAGACATTTCTGCTTCTTAGTCTTA	3643
Db	3541	TATAGAG	TTTACCTTTAATTTTTTAGAGATGTAAGACATTTCTGCTTCTTAGTCTTA	3600
QY	3644	CCTAGTCT	GAACATTTTTTATTCATTAAGATTTTTTAATTAATTAATTAATTAATTAATTAAT	3703
Db	3601	CCTAGTCT	GAACATTTTTTATTCATTAAGATTTTTTAATTAATTAATTAATTAATTAATTAAT	3660
QY	3704	AA	3705	
Db	3661	AA	3662	

AF198454	Homo sapiens epithelial protein lost in neoplasm beta (EPLIN) mRNA, complete cds.	3655 bp	mRNA	linear	PRI 10-JAN-2000
AF198454	AF198454.1	GI:6685006			
ORGANISM	Homo sapiens (human)				
AUTHORS	Maul,R.S. and Chang,D.D.				
TITLE	EPLIN, epithelial protein lost in neoplasm				
JOURNAL	Oncogene 18 (54), 7838-7841 (1999)				
MEDLINE	20087188				
PUBMED	10618726				
REFERENCE	2 (bases 1 to 3655)				
AUTHORS	Maul,R.S. and Chang,D.D.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-OCT-1999) Medicine, UCLA, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA				
FEATURES	Location/Qualifiers				
Source	1..3655				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/chromosome="12"				
	/map="12q13"				
gene	1..3655				
CDS	102..2381				
	/gene="EPLIN"				
	/codon_start=1				
	/product="epithelial protein lost in neoplasm beta"				
	/protein_id="AAF23755.1"				
	/db_xref="GI:6685007"				
	/translation="MESSFPNRRQWTSLSRVTAKELSLVNKNKSSAIVEIFS KYQVA AEETNMKKRNTENLSQHRKGTTLVLKKWENPGLGAESHTDSLNSNSTEIRHAD HPPEAVTSHAAGAKADQEQIHRRLRSPPEALVOGRYPHIKDGEDLKHSTESKIK MENCLESSEHVEKSPISENTDASGKIEKYNVPLNRLKMWFEKGPTOTKIIRAOSR ASGRKISNYSYLDLLEIGPOLSSSTFDEKNESSRNLELRLSETSIKDMAXYQA AVSKQSSNTYNELKASGGEYKHKMEKENVPQPEVCIHQBEKELISAMENSLIAY RSTPAEDSDRSQVKSEVOQPHKPLSPDRASSLSESPPKMKKFKQAPARETCV KQTVYPMERLLANQQVFHISCFRCYCNNKLSLGTASLHGRIYCKPHFNQLFKSG NYDEGFGRPHDKLWASKNENIEILERPQALANARETHSPGVEAPIAKVGVLAASM EKAKSQOBEKDKPABTKLRLAWPPTLGGSGSALESGIKWSPKWPPEDEISKPE VPEDVDLKKLRKRSLSRSPFTVAASFQSTSVKSPKTVSPPIRKGNWSQSESE SVGVRVBAERKQVENAKAKRNGVGTITWONKSEKGTOKRSGKEGHSLEMEINLEN SVSEEDQKRNRYTDEDEDEE"				
	polyA_signal	3632..3637			
BASE COUNT	1197 a 737 c 810 g 911 t				
ORIGIN					
	Quality Match	98.3%	Score 3643.8;	DB 9;	Length 3655;
	Best Local Similarity	99.9%;	Pred. No. 0;		
	Matches 3645;	Conservative	0;	Mismatches	2;
				Indels	0;
				Gaps	0;
Qy	44	GCCTAGGTAGACGCCCGGACCTGTGACAGGCGCTGGTAGCAGCGCAGAGAAAGCGCG	103		
Db	9	GCCTAGGTAGACGCCCGGACCTGTGACAGGCGCTGGTAGCAGCGCAGAGAAAGCGCG	68		
Qy	104	TTTATGCCAGGATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCTCAATTATAGACGG	163		
Db	69	TTTATGCCAGGATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCTCAATTATAGACGG	128		
Qy	164	CAATGGACCTCACTATCATTTAGGGTAAACAGGCCAAAGAACTTTCTCTGTTCACAGAAG	223		
Db	129	CAATGGACCTCACTATCATTTAGGGTAAACAGGCCAAAGAACTTTCTCTGTTCACAGAAG	188		
Qy	224	AGTCAATCGGTATTGTGAAATATTCTCCAGTACCAGAAAGCAGCTGAAGAAACAAAC	283		

Db 1269 TGGGTGGAATGTGAGAAACAGTCTATCCAAATGGAGCGTCTCTTGGCCAAACAGCAGGTG 1328
Qy 1364 TTTCACATCAGCTGCTTCGCTTGCCTCTATGCAACAAACAACTCAGTCTAGAACATAT 1423
Db 1329 TTTCACATCAGCTGCTTCGCTTGCCTCTATGCAACAAACAACTCAGTCTAGAACATAT 1388
Qy 1424 GCATCTTTACATGGAAGAATCTATTGTGAAGCTCCTCAATCAACTCTTTAAATCTAAG 1483
Db 1389 GCATCTTTACATGGAAGAATCTATTGTGAAGCTCCTCAATCAACTCTTTAAATCTAAG 1448
Qy 1484 GGCNACTATGATGAAGCTTTGGGCACAGACACACAGGATCTATGGCAAGCAAAAAT 1543
Db 1449 GGCNACTATGATGAAGCTTTGGGCACAGACACACAGGATCTATGGCAAGCAAAAAT 1508
Qy 1544 GAAACGGAAGATTTTGAGAGACAGCCAGCTTGCAATGCAAGGGAGACCCCTCAC 1603
Db 1509 GAAACGGAAGATTTTGAGAGACAGCCAGCTTGCAATGCAAGGGAGACCCCTCAC 1568
Qy 1604 AGCCAGGGGTAGAAGATGCCCTATTGCTAAGTGGGTGCTCGCTGCAAGTATGGAA 1663
Db 1569 AGCCAGGGGTAGAAGATGCCCTATTGCTAAGTGGGTGCTCGCTGCAAGTATGGAA 1628
Qy 1664 GCAAGGCCCTCTCTCAGCAGAGAGAGGAAGCAAGCCAGCTGAAACCAAGAAAGCTGAGG 1723
Db 1629 GCAAGGCCCTCTCTCAGCAGAGAGAGGAAGCAAGCCAGCTGAAACCAAGAAAGCTGAGG 1688
Qy 1724 ATCGCTGCCACCCCCACCTCAACTTGGAACTTCAGGAAGTTCAGGAAGTTCAGGAAGGATC 1783
Db 1689 ATCGCTGCCACCCCCACCTCAACTTGGAACTTCAGGAAGTTCAGGAAGTTCAGGAAGGATC 1748
Qy 1784 AAAATGTCAAAGCCCAATGGCCCTCTGAAAGCAAGAAATCAGCAAGCCCGAAGTTCCTGAG 1843
Db 1749 AAAATGTCAAAGCCCAATGGCCCTCTGAAAGCAAGAAATCAGCAAGCCCGAAGTTCCTGAG 1808
Qy 1844 GATGTGATCTAGATCTGAAGAAAGTAAAGCAATCTTCTTCACTGAAGAAAGAAAGCCGC 1903
Db 1809 GATGTGATCTAGATCTGAAGAAAGTAAAGCAATCTTCTTCACTGAAGAAAGAAAGCCGC 1868
Qy 1904 CCATTCACTGTAGCAGCTTCACTTTCAAAGCACCCTGTCAAGAGCCCAAAAACCTGTGTCC 1963
Db 1869 CCATTCACTGTAGCAGCTTCACTTTCAAAGCACCCTGTCAAGAGCCCAAAAACCTGTGTCC 1928
Qy 1964 CCACCTATCAGAAAGGCTGGAGCATGTGAGAGAGAGTGAAGAGTCTGTGGTGGAGA 2023
Db 1929 CCACCTATCAGAAAGGCTGGAGCATGTGAGAGAGAGTGAAGAGTCTGTGGTGGAGA 1988
Qy 2024 GTTGAGAAAGGAAACAAAGTGGAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTGGGA 2083
Db 1989 GTTGAGAAAGGAAACAAAGTGGAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTGGGA 2048
Qy 2084 AAAACAACCTGGCAAAACAAAGAAATCTTAAAGGAGAGACAGGAAAGTAAGGAAGGT 2143
Db 2049 AAAACAACCTGGCAAAACAAAGAAATCTTAAAGGAGAGACAGGAAAGTAAGGAAGGT 2108
Qy 2144 CATAGTTTGGAGATGGGAATGAGNAATCTTGTAGAAATGCTGAGACTCCGATGAAGAT 2203
Db 2109 CATAGTTTGGAGATGGGAATGAGNAATCTTGTAGAAATGCTGAGACTCCGATGAAGAT 2168
Qy 2204 GATAACAGCTTCCTCAAAACAAATCTCCAAAGAACCCAGTCTCTGAAATGGTCCAGT 2263
Db 2169 GATAACAGCTTCCTCAAAACAAATCTCCAAAGAACCCAGTCTCTGAAATGGTCCAGT 2228
Qy 2264 TTTGTAGAACACACTTTGCTGAAGAATTCCTACTCAGAAATCAGAAATCCCAGGATGTG 2323
Db 2229 TTTGTAGAACACACTTTGCTGAAGAATTCCTACTCAGAAATCAGAAATCCCAGGATGTG 2288
Qy 2324 GAACTCTGGGAGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAAT 2383
Db 2289 GAACTCTGGGAGGAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAAT 2348
Qy 2384 CGGTATTATGATGAGGATGAGGATGAAGAGTGAAGATTCGAATGATGCTGGGCCCTTAA 2443
Db 2349 CGGTATTATGATGAGGATGAGGATGAAGAGTGAAGATTCGAATGATGCTGGGCCCTTAA 2408

Qy 2444 TTCATGTTAGTGTAGCGAGCCACTGCGCTTTGTCAAAATGTGTATGCACATAACGACGTA 2503
Db 2409 TTCATGTTAGTGTAGCGAGCCACTGCGCTTTGTCAAAATGTGTATGCACATAACGACGTA 2468
Qy 2504 TCCAGCATGAAATGTAAATTTACTTTGGAAGTAACTTTGGAAAGAAATTCCTTCTTAAAT 2563
Db 2469 TCCAGCATGAAATGTAAATTTACTTTGGAAGTAACTTTGGAAAGAAATTCCTTCTTAAAT 2528
Qy 2564 CAAAAACAAAAACAAAAACAAAAACAACTTCTAAATACTAGAGATAACTTTACTTA 2623
Db 2529 CAAAAACAAAAACAAAAACAAAAACAACTTCTAAATACTAGAGATAACTTTACTTA 2588
Qy 2624 AATTCTTCATTTTAGCAGTGAATATGCATATGCTGTAAAGCTTTGTAACTGGGAAA 2683
Db 2589 AATTCTTCATTTTAGCAGTGAATATGCATATGCTGTAAAGCTTTGTAACTGGGAAA 2648
Qy 2684 TATTCCACCTGATAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAGGTAGA 2743
Db 2649 TATTCCACCTGATAATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAGGTAGA 2708
Qy 2744 TAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGACATACAGAGAA 2803
Db 2709 TAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGACATACAGAGAA 2768
Qy 2804 TTTAGGGCTTAAACATTTACGACTGAATGCACCTTTAGTATAAAGGGCAGCTTTGTATAT 2863
Db 2769 TTTAGGGCTTAAACATTTACGACTGAATGCACCTTTAGTATAAAGGGCAGCTTTGTATAT 2828
Qy 2864 TTTTAAATGAATACCAATTTAAATTTTGTATTTTACCTGTTAAGAGATTTATTAGTCTT 2923
Db 2829 TTTTAAATGAATACCAATTTAAATTTTGTATTTTACCTGTTAAGAGATTTATTAGTCTT 2888
Qy 2924 TAAATTTTGTAGTAAATTTTCTTGTCTGTGATATATATAGGAAATTTACTACTTTATGTC 2983
Db 2889 TAAATTTTGTAGTAAATTTTCTTGTCTGTGATATATATAGGAAATTTACTACTTTATGTC 2948
Qy 2984 CTGCTCTCTAAACTACATCTGAACCTCGAGCTCCTCGAGTATATACACAGAGCACTTT 3043
Db 2949 CTGCTCTCTAAACTACATCTGAACCTCGAGCTCCTCGAGTATATACACAGAGCACTTT 3008
Qy 3044 TTGAGCAATTTGAAACCAACCACTACACTCTTCGGTGCTTAGAGAGATCTGCTGTCTCC 3103
Db 3009 TTGAGCAATTTGAAACCAACCACTACACTCTTCGGTGCTTAGAGAGATCTGCTGTCTCC 3068
Qy 3104 AAATAAGCTTTGTATCTGCCAGTGAATTTACTGTACTCCAA/TGATTTGCTTTCTTCT 3163
Db 3069 AAATAAGCTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTTCT 3128
Qy 3164 GGTGATATCTGTCTCTCATTAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGG 3223
Db 3129 GGTGATATCTGTCTCTCATTAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGG 3188
Qy 3224 ATCACTGTCCCCCATCTTCCGTGTAGACCAAGTGAAGAGTTTAAAGGAGGAAGAA 3283
Db 3189 ATCACTGTCCCCCATCTTCCGTGTAGACCAAGTGAAGAGTTTAAAGGAGGAAGAA 3248
Qy 3284 AGAATCTTTACACACTTTAGCTCAGACCTCTAAACCCCTGTATTTCCCTTATGATGTC 3343
Db 3249 AGAATCTTTACACACTTTAGCTCAGACCTCTAAACCCCTGTATTTCCCTTATGATGTC 3308
Qy 3344 CCGTTTTGAGACACTAAATTTTAAATACTTACTAGCTCTGAAATATATTTGATTTTATC 3403
Db 3309 CCGTTTTGAGACACTAAATTTTAAATACTTACTAGCTCTGAAATATATTTGATTTTATC 3368
Qy 3404 ACAGTATTTCTAGGGTGAATTTAAACCACTATAGCCCTTTTCTTGGGATCATTTCTA 3463
Db 3369 ACAGTATTTCTAGGGTGAATTTAAACCACTATAGCCCTTTTCTTGGGATCATTTCTA 3428
Qy 3464 GTCTTAAAGTTTGGGCACTTAAACCTTGAGTACATTTGTTGTACACAGTTGATTTCC 3523
Db 3429 GTCTTAAAGTTTGGGCACTTAAACCTTGAGTACATTTGTTGTACACAGTTGATTTCC 3488

QY	1439	AGAATCTAATTGTAAGGCTCACTTCAATCAACTCTTTAAATCTTAAGGCAACTATGATGA	1498
Db	1299	AGAATCTAATTGTAAGGCTCACTTCAATCAACTCTTTAAATCTTAAGGCAACTATGATGA	1358
QY	1499	GGCTTTGGGCACAGACACACAAGGATCTATGGSCAAGCAAAATGAAACGAAGAGATT	1558
Db	1359	GGCTTTGGGCACAGACACACAAGGATCTATGGSCAAGCAAAATGAAACGAAGAGATT	1418
QY	1559	TTGGAGAGACCCAGCTTGCAAAATGCNAGGAGACCCCTCACAGCCAGGGGTAGAA	1618
Db	1419	TTGGAGAGACCCAGCTTGCAAAATGCNAGGAGACCCCTCACAGCCAGGGGTAGAA	1478
QY	1619	GATGCCCCCTTCTAAGGTGGGTGCTCGCTGCAAGTATGGAAGCCCAAGCCCTCTCT	1678
Db	1479	GATGCCCCCTTCTAAGGTGGGTGCTCGCTGCAAGTATGGAAGCCCAAGCCCTCTCT	1538
QY	1679	CAGCAGAGAGGAAGAACAGCCAGCTGAACCAAGAGCTGAGGATCGCCTGGCCACCC	1738
Db	1539	CAGCAGAGAGGAAGAACAGCCAGCTGAACCAAGAGCTGAGGATCGCCTGGCCACCC	1598
QY	1739	CCCACTGAACCTTGGAGTTTCAGGAAGTGCCTTGGAGGAAGGATCAAAATGTCAAAGCCC	1798
Db	1599	CCCACTGAACCTTGGAGTTTCAGGAAGTGCCTTGGAGGAAGGATCAAAATGTCAAAGCCC	1658
QY	1799	AAATGGCCTCTGAAGCAGAAATCAGACGCCCGAAGTTCCCTGAGGATGTCATCTAGAT	1858
Db	1659	AAATGGCCTCTGAAGCAGAAATCAGACGCCCGAAGTTCCCTGAGGATGTCATCTAGAT	1718
QY	1859	CTGAAGAAAGCTAAGACGATCTCTTCACTGAAGGAAGAGCGCCCATCTCACTGTAGCA	1918
Db	1719	CTGAAGAAAGCTAAGACGATCTCTTCACTGAAGGAAGAGCGCCCATCTCACTGTAGCA	1778
QY	1919	GCTTCAATTTCAAAGCACTCTGTCAAGACCCCAAAAACCTGTGCCCACTATCAGGAAA	1978
Db	1779	GCTTCAATTTCAAAGCACTCTGTCAAGACCCCAAAAACCTGTGCCCACTATCAGGAAA	1838
QY	1979	GGCTGAGAGATGTCAGACGAGAGTGAAGAGTCTGTGGTGGAGAGTTCGCAAAAGGAAA	2038
Db	1839	GGCTGAGAGATGTCAGACGAGAGTGAAGAGTCTGTGGTGGAGAGTTCGCAAAAGGAAA	1898
QY	2039	CAAGTGGAAAAATGCCAAGGCTTCTAAGAGAATGGGAATGTGGAAAAACACCTGGCAA	2098
Db	1899	CAAGTGGAAAAATGCCAAGGCTTCTAAGAGAATGGGAATGTGGAAAAACACCTGGCAA	1958
QY	2099	AACAAAGAAATCTAAAGGAGAGACAGGGAAGAGAAAGTAAAGGAAGTCAATAGTTTGAGAGTG	2158
Db	1959	AACAAAGAAATCTAAAGGAGAGACAGGGAAGAGAAAGTAAAGGAAGTCAATAGTTTGAGAGTG	2018
QY	2159	GAGAATGAGAATCTTGTAGAAAATGGTCAGACTCCGATGAAGATGATAACAGCTTCCTC	2218
Db	2019	GAGAATGAGAATCTTGTAGAAAATGGTCAGACTCCGATGAAGATGATAACAGCTTCCTC	2078
QY	2219	AAACAAATCTCCCAAGAACCCAGTCTCTGAATGGTCAGTTTGTAGACAAACACC	2278
Db	2079	AAACAAATCTCCCAAGAACCCAGTCTCTGAATGGTCAGTTTGTAGACAAACACC	2138
QY	2279	TTTGTCTGAAGAAATCTCACTACTCAGAAATCCAGGATGTGGAATCTCTGGAGGGA	2338
Db	2139	TTTGTCTGAAGAAATCTCACTACTCAGAAATCCAGGATGTGGAATCTCTGGAGGGA	2198
QY	2339	GAAGTGGTCAAGAGAGTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAG	2398
Db	2199	GAAGTGGTCAAGAGAGTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAG	2258
QY	2399	GATGAGGATGAAGAGTGAAGAAATTCGAATGATGCTGGGCCCTTAAATTCATGTTAGTGTA	2458
Db	2259	GATGAGGATGAAGAGTGAAGAAATTCGAATGATGCTGGGCCCTTAAATTCATGTTAGTGTA	2318
QY	2459	GGAGGCACTGCCCTTTGTCAAAATGTGATGCACATAGCAGGTATCCACGATGAATG	2518
Db	2319	GGAGGCACTGCCCTTTGTCAAAATGTGATGCACATAGCAGGTATCCACGATGAATG	2378

QY	2519	TAATTTACTTGGAAAGTAACTTTGGAAAAAGAAATTCCTTCTTAAAAATCAAAAAACAAACAA	2578
Db	2379	TAATTTACTTGGAAAGTAACTTTGGAAAAAGAAATTCCTTCTTAAAAATCAAAAAACAAACAA	2438
QY	2579	AAAAACAAAAAACAACACTTCTAAATACTAGAGATAACTTTTACTTTTADATTTCTTATTAG	2638
Db	2439	AAAAACAAAAAACAACACTTCTAAATACTAGAGATAACTTTTACTTTTAAATTTCTTATTAG	2498
QY	2639	CAGTGATGATATGCATAGTCTGTAAAGCTTGTAACTTGGGAAATATTTCCACCTGTATAA	2698
Db	2499	CAGTGATGATATGCATAGTCTGTAAAGCTTGTAACTTGGGAAATATTTCCACCTGTATAA	2558
QY	2699	TAGCCCAAGTTCTACTGTATTTCCCAAGGCAATATTAAGGTAGATAGATAGTAGTAGT	2758
Db	2559	TAGCCCAAGTTCTACTGTATTTCCCAAGGCAATATTAAGGTAGATAGATAGTAGTAGT	2618
QY	2759	ATATTTGTACACACTATTTTGGAAATAGAGAACTACAGAGGAAATTTAGGGGCTTAAAC	2818
Db	2619	ATATTTGTACACACTATTTTGGAAATAGAGAACTACAGAGGAAATTTAGGGGCTTAAAC	2678
QY	2819	ATTACGACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATATTTTAAATGAATACC	2878
Db	2679	ATTACGACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATATTTTAAATGAATACC	2738
QY	2879	AATTTAATTTTTTGTATTTTACCTGTGTAAAGAGATATTTAGTCTTTTAAATTTTTTAGGTT	2938
Db	2739	AATTTAATTTTTTGTATTTTACCTGTGTAAAGAGATATTTAGTCTTTTAAATTTTTTAGGTT	2798
QY	2939	AATTTCTTGTGTGATATATAGAGAAATTTACTACTTTTATGCTCTCTAAACTA	2998
Db	2799	AATTTCTTGTGTGATATATAGAGAAATTTACTACTTTTATGCTCTCTAAACTA	2858
QY	2999	CATCTGAACCTCGACGCTCTGAGGTATAACACAGAGCACTTTTGGAGCAATTTGAAA	3058
Db	2859	CATCTGAACCTCGACGCTCTGAGGTATAACACAGAGCACTTTTGGAGCAATTTGAAA	2918
QY	3059	AACCAACCTACACTCTTTCGGTGCTTAGAGAGATCTGCTGTCTCCCAATTAAGCTTTTGT	3118
Db	2919	AACCAACCTACACTCTTTCGGTGCTTAGAGAGATCTGCTGTCTCCCAATTAAGCTTTTGT	2978
QY	3119	TCTGCCAGTGAATTTACTGTACTCCAAATGATGCTTTCTTTCTGGTGATATCTGTCT	3178
Db	2979	TCTGCCAGTGAATTTACTGTACTCCAAATGATGCTTTCTTTCTGGTGATATCTGTCT	3038
QY	3179	TCTCATATTTACTGAAAGCTGCATATTTTAGTAAATACCTTCGGGATCACTGTCCCCAT	3238
Db	3039	TCTCATATTTACTGAAAGCTGCATATTTTAGTAAATACCTTCGGGATCACTGTCCCCAT	3098
QY	3239	CTTCCGTGTTAGAGCAAGTGAAGGTTTTAAAGGAGGAAGAAAGAACTGTCTTACAC	3298
Db	3099	CTTCCGTGTTAGAGCAAGTGAAGGTTTTAAAGGAGGAAGAAAGAACTGTCTTACAC	3158
QY	3299	CACCTTGAGCTCAGACCTCTAAACCCCTGTATTTCCCTTTATGATGTCCCTTTTGGAGAC	3358
Db	3159	CACCTTGAGCTCAGACCTCTAAACCCCTGTATTTCCCTTTATGATGTCCCTTTTGGAGAC	3318
QY	3359	TAATTTTAAATCTACTAGCTCTGAAATATATTTAGTATTTTATCAGATTTCTCAGGG	3418
Db	3219	TAATTTTAAATCTACTAGCTCTGAAATATATTTAGTATTTTATCAGATTTCTCAGGG	3278
QY	3419	TGAATTTAAACCAACTATAGCCCTTTTCTGGGATGATTTTCTAGTCTTTAAGTTTGGG	3478
Db	3279	TGAATTTAAACCAACTATAGCCCTTTTCTGGGATGATTTTCTAGTCTTTAAGTTTGGG	3338
QY	3479	GACATTATAAAGTGTAGTACATTTTGTACAGTTGATATTTCCAAATTTGTATGGATGG	3538
Db	3339	GACATTATAAAGTGTAGTACATTTTGTGTACAGTTGATATTTCCAAATTTGTATGGATGG	3398
QY	3539	GAGGAGAGTGCTTTAAGCTGTAGGCTTTTCTTTGTACTGCAATTTTATAGAGATTTAGCT	3598
Db	3399	GAGGAGAGTGCTTTAAGCTGTAGGCTTTTCTTTGTACTGCAATTTTATAGAGATTTAGCT	3458
QY	3599	TTAATATTTTTAGAGATGTAAAAACATTTCTGCTTTCTTAGTCTTACCTAGTCTGAAACAT	3658

Db	3459	TTAATATTTTGTAGAGTGAACACATCTGCTTTCTTAGCTTACCTAGTCTGAAACAT	3518
Qy	3659	TTTATTCAATAAGATTAAATTAATAATTG	3690
Db	3519	TTTATTCAATAAGATTAAATTAATAATTG	3550
RESULT 5			
LOCUS	AK000335		
DEFINITION	Homo sapiens cDNA FLJ20328 fis, clone HEP10039.	linear	PRI 22-FEB-2000
ACCESSION	AK000335		
VERSION	AK000335.1	GI:7020350	
KEYWORDS	oligo capping, fis (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (sites)		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,		
	Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,		
	Suzuki,Y., Obayashi,M., Nishi,T., Shibahata,T., Tanaka,T.,		
	Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 3277)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,		
	Shibahara,T., Tanaka,T. and Nakamura,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,		
	University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1,		
	Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ms.u-tokyo.ac.jp,		
	Tel:81-3-5449-5286, Fax:81-3-5449-5416)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of		
	International Trade and Industry of Japan; cDNA full insert		
	sequencing; Research Association for Biotechnology; cDNA library		
	construction, 5'- & 3'-end one pass sequencing; Department of		
	Virology and Human Genome Center, Institute of Medical Science,		
	University of Tokyo (partly supported by Science and Technology		
	Agency).		
FEATURES	Location/Qualifiers		
source	1..3277		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="HEP10039"		
	/cell_line="RepG2"		
	/cell_type="Hepatoma"		
	/clone_lib="HEP"		
	/note="cloning vector pME18SFL3"		
	179..1978		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="BAA91092.1"		
	/db_xref="GI:7020351"		
	/translation="MENCLEGRHEVEKSEISNTDASGKIEKNVPLNRLKWPFEK		
	EPQTQKILRAQSRSASGRKISNSYSLDLEIGPGLSSSTPSEKNSRNLELRL		
	SEISIKDRMAKYQAAYSKQSSNTYNNELKASGGEIKHMEQENVPFGEVCIHQ		
	EGKISANENSLAVRSTPADDSDRQVSKSEVQVHPKPLSPDSRSLSSESPPKA		
	MKXPAPARTCEQKTVPMERLLANQVPHISCFRCSCYNNKLSLGTYSALHGR1		
	YCKPHEQLFKSKNGYDEGFGRPHKDLMAKNENEIILERPAOLANARETPHSGVE		
	DAP1AKVGLVAAMEAKSSQEKDKPAETKKLR1AWPPPTFLSGSGLSEGIKMS		
	KPKWPPDETSKPEVPDVLDDLKLRSSSLKERSRPTVAASFQSTSVKSPKIVSP		
	PIKRGWMSQSESVEGRVAERKQVENAKSKNGNVGKTTWNKESKGETGRKSE		
	QDVLMEENLVNGADSDSDSDNLFUKQSPQEPKSLNWSFFDNTFAEFTTQKQS		
	QVLEMEGEYVVEEIEQIKRNYDEDEDEE"		
BASE COUNT	1089 a	651 c	703 g
ORIGIN		834 t	
Query Match	88.08;	Score 3262.2;	DB 9; Length 3277;
Best Local Similarity	99.98;	Pred. No. 0;	
Matches 3264; Conservative	0; Mismatches	3; Indels	0; Gaps

QY 1519 CAAGGATCTATTGGCAAGCAAAAATGAAAA CGAAGAGATTTTGGAGAGACAGCCAGCT 1578
DB 1081 CAAGGATCTATTGGCAAGCAAAAATGAAAA CGAAGAGATTTTGGAGAGACAGCCAGCT 1140
QY 1579 TGCMAATGCAAGGAGACCCCTCACAGCCCGAGGGGTAGAGATGCCCCCTAATCTAAGGT 1638
DB 1141 TGCMAATGCAAGGAGACCCCTCACAGCCCGAGGGGTAGAGATGCCCCCTAATCTAAGGT 1200
QY 1639 GGGTGTCTCGCTGCAAGTATGGAAGCCAGAGCTCTCTCTCAGCAGAGAGAGAAACAA 1698
DB 1201 GGGTGTCTCGCTGCAAGTATGGAAGCCAGAGCTCTCTCTCAGCAGAGAGAGAAACAA 1260
QY 1699 GCCAGCTGAAACCAAGAGCTGAGGATGCGCTGGCCACCCGCCACTGAACTTGGAAAGTTC 1758
DB 1261 GCCAGCTGAAACCAAGAGCTGAGGATGCGCTGGCCACCCGCCACTGAACTTGGAAAGTTC 1320
QY 1759 AGGAAGTGCCTTGGAGAAAGGATCAAAATGTCAAAGCCCAAAATGCCCTCTGAAAGCA 1818
DB 1321 AGGAAGTGCCTTGGAGAAAGGATCAAAATGTCAAAGCCCAAAATGCCCTCTGAAAGCA 1380
QY 1819 AATCAGCAAGCCGAAAGTCTCTGAGGATGTCGATCTAGATCTGAAAGAGCTAAGACGATC 1878
DB 1381 AATCAGCAAGCCGAAAGTCTCTGAGGATGTCGATCTAGATCTGAAAGAGCTAAGACGATC 1440
QY 1879 TTCTTCACTGAAGAAAGAGCCGCCATTCAGTGTAGCAGCTTCAATTTCAAAGCACCTC 1938
DB 1441 TTCTTCACTGAAGAAAGAGCCGCCATTCAGTGTAGCAGCTTCAATTTCAAAGCACCTC 1500
QY 1939 TGTCAAGAGCCCAAAAACCTGTGTCCCACCTATCAGGAAAGGCTGGAGCATGTCAAGCA 1998
DB 1501 TGTCAAGAGCCCAAAAACCTGTGTCCCACCTATCAGGAAAGGCTGGAGCATGTCAAGCA 1560
QY 1999 GAGTGAAGAGTCTGTGGTGGAGAGAGTGTGAGAAAGAAACAAAGTGGAAAAATGCCAAGC 2058
DB 1561 GAGTGAAGAGTCTGTGGTGGAGAGAGTGTGAGAAAGAAACAAAGTGGAAAAATGCCAAGC 1620
QY 2059 TTCTAAGAGAAATGGGAATGTGGGAAAAACAACTGGCAAAAACAAAGATCTAAAGGAGA 2118
DB 1621 TTCTAAGAGAAATGGGAATGTGGGAAAAACAACTGGCAAAAACAAAGATCTAAAGGAGA 1680
QY 2119 GACAGGAAAGAGAGATGAAGAAAGTCTATAGTTTGGAGATGGAGATGAAGATCTTGTAGA 2178
DB 1681 GACAGGAAAGAGAGATGAAGAAAGTCTATAGTTTGGAGATGGAGATGAAGATCTTGTAGA 1740
QY 2179 AATGTGTGAGATCTCGATGAAGATGATTAACAGCTTCTCTCAAAACAAATCTCCACAAGA 2238
DB 1741 AATGTGTGAGATCTCGATGAAGATGATTAACAGCTTCTCTCAAAACAAATCTCCACAAGA 1800
QY 2239 ACCCAAGTCTCTCAATGTGTGAGTTTGTAGACAACACCTTTGTGAAAGAAATTCACATC 2298
DB 1801 ACCCAAGTCTCTCAATGTGTGAGTTTGTAGACAACACCTTTGTGAAAGAAATTCACATC 1860
QY 2299 TCAGAATCAGAAATCCAGGATGTGGAATCTCTGGAGGGAGAGTGGTCAAGAGCTCTC 2358
DB 1861 TCAGAATCAGAAATCCAGGATGTGGAATCTCTGGAGGGAGAGTGGTCAAGAGCTCTC 1920
QY 2359 TGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGACA 2418
DB 1921 TGTGGAAGAACAGATAAAGAGAAATCGGTATTATGATGAGGATGAGGATGAAGAGTGACA 1980
QY 2419 AATTGCAATGATGCTGGGCTTAAATTCATGTATTGTAGGAGCCACTGCCCCCTTTGTC 2478
DB 1981 AATTGCAATGATGCTGGGCTTAAATTCATGTATTGTAGGAGCCACTGCCCCCTTTGTC 2040
QY 2479 AAAATGTGATGCACATGAAGCAGGTATCCAGCATGAAATGTAATTTACTTGGAAAGTAACT 2538
DB 2041 AAAATGTGATGCACATGAAGCAGGTATCCAGCATGAAATGTAATTTACTTGGAAAGTAACT 2100
QY 2539 TTGGAAAAGAAATTCCTTCTTAAAAATCAAAAAACAAAAACAAAAACAAAAACAAATTC 2598
DB 2101 TTGGAAAAGAAATTCCTTCTTAAAAATCAAAAAACAAAAACAAAAACAAAAACAAATTC 2160
QY 2599 TAAATCTAGAGATAACTTTTACTTAAATTCCTTAAATTTAGCAGTGTATGCATATAAGT 2658

DB 2161 TAAATCTAGAGATAACTTTTACTTAAATTCCTTAAATTTAGCAGTGTATATGCATATAAGT 2220
QY 2659 GCTGTAAAGCTTTTAACCTGGGAAAATATTCACCTGTATAATAGCCAGATTTCTACTGTAT 2718
DB 2221 GCTGTAAAGCTTTTAACCTGGGAAAATATTCACCTGTATAATAGCCAGATTTCTACTGTAT 2280
QY 2719 TCCCAAAAAGGCAATATTAAAGGTAGATAGATATTAGTAGTATATTGTTTACACACTATTTT 2778
DB 2281 TCCCAAAAAGGCAATATTAAAGGTAGACAGATGATTAGTAGTATATTGTTTACACACTATTTT 2340
QY 2779 GGAATTTAGAGAACATACACAGAGGAATTTAGGGCTTTAAACATTACGACTGAATGCACATT 2838
DB 2341 GGAATTTAGAGAACATACACAGAGGAATTTAGGGCTTTAAACATTACGACTGAATGCACATT 2400
QY 2839 AGTATAAAAGGCAACAGTTTGTATATTTTAAATGAATACCAATTTAAATTTTGTATTTT 2898
DB 2401 AGTATAAAAGGCAACAGTTTGTATATTTTAAATGAATACCAATTTAAATTTTGTATTTT 2460
QY 2899 ACCTGTTAAGAGATATTATTAGTCTTTTAAATTTTGTAGTTTAAATTTTGTGTGTATATA 2958
DB 2461 ACCTGTTAAGAGATATTATTAGTCTTTTAAATTTTGTAGTTTAAATTTTGTGTGTATATA 2520
QY 2959 TATGAGGAATTTACTACTTTTATGCTGCTCTCTAAACTACATCCTGAACTCGAGCTCCT 3018
DB 2521 TATGAGGAATTTACTACTTTTATGCTGCTCTCTAAACTACATCCTGAACTCGAGCTCCT 2580
QY 3019 GAGGTATATAACACAGAGCACTTTTGGAGCAATTTGAGGCAATTTGAAACCAACCTACACTCTTCGG 3078
DB 2581 GAGGTATATAACACAGAGCACTTTTGGAGCAATTTGAGGCAATTTGAAACCAACCTACACTCTTCGG 2640
QY 3079 TGTTAGAGAGATCTGCTGCTCTCCCAATAAGCTTTTGTATCTGCGAGTGAATTTACTGT 3138
DB 2641 TGTTAGAGAGATCTGCTGCTCTCCCAATAAGCTTTTGTATCTGCGAGTGAATTTACTGT 2700
QY 3139 ACTCAAAATGATGCTTTTCTGTTGATATCTGTGCTTCTCATAAATTAAGTAAAGCT 3198
DB 2701 ACTCAAAATGATGCTTTTCTGTTGATATCTGTGCTTCTCATAAATTAAGTAAAGCT 2760
QY 3199 GCAATATTTTAGTAATACCTTCGGGATCAGTGTCCCCCATCTTCGGTGTAGAGCAAAAGT 3258
DB 2761 GCAATATTTTAGTAATACCTTCGGGATCAGTGTCCCCCATCTTCGGTGTAGAGCAAAAGT 2820
QY 3259 GAAGAGTTTAAAGGAGGAAGAAAGAACTGTCTTACACCACTTGAGCTCAGACCTCTA 3318
DB 2821 GAAGAGTTTAAAGGAGGAAGAAAGAACTGTCTTACACCACTTGAGCTCAGACCTCTA 2880
QY 3319 AACCTGTATTTCCCTTATGATGTCCCTTTTTCAGACACTAAATTTTAAATACTTACTA 3378
DB 2881 AACCTGTATTTCCCTTATGATGTCCCTTTTTCAGACACTAAATTTTAAATACTTACTA 2940
QY 3379 GCTCTGAAATATATTGATTTTATCAAGATTTCTCAGGTGAAATTTAAACCAACTATAG 3438
DB 2941 GCTCTGAAATATATTGATTTTATCAAGATTTCTCAGGTGAAATTTAAACCAACTATAG 3000
QY 3439 GCCCTTTTCTGGATGATTTTCTAGTCTTAAAGTTTGGGACATTTATAAAGTGTAGTAC 3498
DB 3001 GCCCTTTTCTGGATGATTTTCTAGTCTTAAAGTTTGGGACATTTATAAAGTGTAGTAC 3060
QY 3499 ATTTGTTGTACACAGTTGATATTCCAAATTTGTATGGATGGAGGGAGAGTGTCTTAAGC 3558
DB 3061 ATTTGTTGTACACAGTTGATATTCCAAATTTGTATGGATGGAGGGAGAGTGTCTTAAGC 3120
QY 3559 TGTAGGCTTTTCTTGTACTGCATTTATAGAGATTTTAGCTTTAATATTTTGTAGAGATGT 3618
DB 3121 TGTAGGCTTTTCTTGTACTGCATTTATAGAGATTTAGCTTTAATATTTTGTAGAGATGT 3180
QY 3619 AAAACATTCCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATTTCAATAAGATTTT 3678
DB 3181 AAAACATTCCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATTTCAATAAGATTTT 3240
QY 3679 AATTAAATTTTGAAGAAAAAAGAAAAA 3705

Db 3241 AATTAAATTTTCAAAAAAAAAAAAAA 3267

RESULT 6
AX017487
LOCUS AX017487 2783 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 31 from Patent WO9947655.
ACCESSION AX017487
VERSION AX017487.1 GI:10042284
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE 1
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pillarsky,C.
TITLE Human nucleic acid sequences from normal breast tissue
JOURNAL Patent: WO 9947655-A 31 23-SEP-1999.
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..2783
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 901 a 526 c 609 g 747 t
ORIGIN

Query Match 72.8%; Score 2698.2; DB 6; Length 2783;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2700; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1001 AATGAGCTGAAGCCAGTGGTGGCGAATCAATAATGGACCAAGGAGAAT 1060
DB 40 AATGAGCTGAAGCCAGTGGTGGCGAATCAATAATGGACCAAGGAGAAT 99
QY 1061 GTGCCCCCAGGTCCTGAGGTCTGCATCACCCTCAGGAAGGGGAAAAGATTTCGCAAAAT 1120
DB 100 GTGCCCCCAGGTCCTGAGGTCTGCATCACCCTCAGGAAGGGGAAAAGATTTCGCAAAAT 159
QY 1121 GAGTAATAGCTGGCAGTCCGTTCCACCCCTCCGAAAGATGACTCCCGTGACTCCAGGTT 1180
DB 160 GAGTAATAGCTGGCAGTCCGTTCCACCCCTCCGAAAGATGACTCCCGTGACTCCAGGTT 219
QY 1181 AAGAGTAGGTTCAACAGCCTGTCCATCCCAAGCCACTAGTCCAGATTCAGAGCCCTCC 1240
DB 220 AAGAGTAGGTTCAACAGCCTGTCCATCCCAAGCCACTAGTCCAGATTCAGAGCCCTCC 279
QY 1241 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 1300
DB 280 AGTCTTTCTGAAAGTTCTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 339
QY 1301 ACCTGCGTGGAAATGTCAAGACAGTCTATCCAAATGGAGCGTCTCTTTGGCCCAACCCAG 1360
DB 340 ACCTGCGTGGAAATGTCAAGACAGTCTATCCAAATGGAGCGTCTCTTTGGCCCAACCCAG 399
QY 1361 GTGTTTCACATCAGCTGCTCCGTTGCTCTCTATTGGAACACAACTCAGTCTAGGAACA 1420
DB 400 GTGTTTCACATCAGCTGCTCCGTTGCTCTCTATTGGAACACAACTCAGTCTAGGAACA 459
QY 1421 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 1480
DB 460 TATGCATCTTTACATGGAAGNAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 519
QY 1481 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCTATGGGCAAGCAAA 1540
DB 520 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCTATGGGCAAGCAAA 579
QY 1541 AATGAAAACGAAGATTTTGGAGAGACCAGCCAGCTTGCAAAATGCAAGGAGACCCCT 1600
DB 580 AATGAAAACGAAGATTTTGGAGAGACCAGCCAGCTTGCAAAATGCAAGGAGACCCCT 639

QY 1601 CACAGCCCAGGGGTAGAAATGCCCTATTGTGAAGTGGGTGCTCTGCTGCAAGATATG 1660
DB 640 CACAGCCCAGGGGTAGAAATGCCCTATTGTGAAGTGGGTGCTCTGCTGCAAGATATG 699
QY 1661 GAAGCCAAAGCCTCTCTCAGCAGGAAGGAAGACAAAGCCAGCTGAAACCAAGAAAGCTG 1720
DB 700 GAAGCCAAAGCCTCTCTCAGCAGGAAGGAAGACAAAGCCAGCTGAAACCAAGAAAGCTG 759
QY 1721 AGGATCGCTGGCCACCCCTCACTGAAGTTGGAAGTTCAGGAAGTCCCTTTGAGGAAGGG 1780
DB 760 AGGATCGCTGGCCACCCCTCACTGAAGTTCAGGAAGTCCCTTTGAGGAAGGG 819
QY 1781 ATCAAAATGTCAAAAGCCCAAAATGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGTTCT 1840
DB 820 ATCAAAATGTCAAAAGCCCAAAATGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGTTCT 879
QY 1841 GAGGATGTGATCTAGATCTGAAGAAGCTTAAGACGATCTTCTTCACTGAAGGAAGAAGC 1900
DB 880 GAGGATGTGATCTAGATCTGAAGAAGCTTAAGACGATCTTCTTCACTGAAGGAAGAAGC 939
QY 1901 CGCCCATTCACGTAGCAGCTTCATTTCAAAGCAGCTCTGTCAGAGGCCCAAAAAGCTG 1960
DB 940 CGCCCATTCACGTAGCAGCTTCATTTCAAAGCAGCTCTGTCAGAGGCCCAAAAAGCTG 999
QY 1961 TCCCCACCTATCAGGAAGGCTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGGTGA 2020
DB 1000 TCCCCACCTATCAGGAAGGCTGGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGGTGA 1059
QY 2021 AGATTGCGAAGGAACCAAGTGGAAATGCCAAGGCTTCTAAGAAGAAATCGGAATGTG 2080
DB 1060 AGATTGCGAAGGAACCAAGTGGAAATGCCAAGGCTTCTAAGAAGAAATGGGAATGTG 1119
QY 2081 GGAAGAAACCACTGGCAAAACAAAGATCTAAGAGAGACAGGGAGAGAAGTAAAGAA 2140
DB 1120 GGAAGAAACCACTGGCAAAACAAAGATCTAAGAGAGACAGGGAGAGAAGTAAAGAA 1179
QY 2141 GGTCTAGTGTGGAGATGAGAAATGAGAAATCTGTAGAAAATGGTGCAGACTCCGATGA 2200
DB 1180 GGTCTAGTGTGGAGATGAGAAATGAGAAATCTGTAGAAAATGGTGCAGACTCCGATGA 1239
QY 2201 GATGATTAACAGCTTCTCTCAACCAACAATCTCCAAGAAACCCCAAGTCTCTGAATGGT 2260
DB 1240 GATGATTAACAGCTTCTCTCAACCAACAATCTCCAAGAAACCCCAAGTCTCTGAATGGT 1299
QY 2261 AGTTTGTGAGACAACTTTGCTGAGAAATTCATCTCAGAAATCAGAAATCCCAGGAT 2320
DB 1300 AGTTTGTGAGACAACTTTGCTGAGAAATTCATCTCAGAAATCAGAAATCCCAGGAT 1359
QY 2321 GTGGAACCTCTGGAGGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGACAGATAAAGAGA 2380
DB 1360 GTGGAACCTCTGGAGGGGAGAGTGGTCAAGAGCTCTCTGTGGAAGACAGATAAAGAGA 1419
QY 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGCACAAATTCGAATGATGCTGGCCCT 2440
DB 1420 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGCACAAATTCGAATGATGCTGGCCCT 1479
QY 2441 AAATTCATGTTAGTGTAGCGAGCCACTGCCCTTTGTCAAAATGTGATGCAATGAAGCAG 2500
DB 1480 AAATTCATGTTAGTGTAGCGAGCCACTGCCCTTTGTCAAAATGTGATGCAATGAAGCAG 1539
QY 2501 GTATCCAGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAAGAAATTCCTTCTTAA 2560
DB 1540 GTATCCAGCATGAAATGTAATTTACTTGGAAAGTAACTTTGGAAAAGAAATTCCTTCTTAA 1599
QY 2561 AATCAAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAC 2620
DB 1600 AATCAAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAA 1659
QY 2621 TTAAATTTCTTCAATTTTAGCAGTGATGATGATGATGATGATGATGATGATGATGATGATG 2680
DB 1660 TTAAATTTCTTCAATTTTAGCAGTGATGATGATGATGATGATGATGATGATGATGATGATG 1719

Qy	2681	AAATATTCCACCTGATAAATAGCCAGAGATTCCTACTGTATTTCCC AAAAGC CAATATTAAAGGT	274
Db	1720	AAATATTCCACCTGATAAATAGCCAGAGATTCCTACTGTATTTCCC AAAAGC CAATATTAAAGGT	1779
Qy	2741	AGATAGATGATTAGTAGTATATTGTTTACACACATATTTTTGGAAATTAGAGAACAATACAGAAG	2800
Db	1780	AGATAGATGATTAGTAGTATATTGTTTACACACATATTTTTGGAAATTAGAGAACAATACAGAAG	1839
Qy	2801	GAATTTAGGGCTTAACACATTACGACCTGAATGCACCTTTAGTATTAAGGGCACAGTTTGTA	2860
Db	1840	GAATTTAGGGCTTAACACATTACGACCTGAATGCACCTTTAGTATTAAGGGCACAGTTTGTA	1899
Qy	2861	TATTTTTAAATAGAAATACCAATTTAAATTTTTTAGTATTTTACCTGTTAAGACATTTATTAGT	2920
Db	1900	TATTTTTAAATAGAAATACCAATTTAAATTTTTTAGTATTTTACCTGTTAAGACATTTATTAGT	1959
Qy	2921	CTTTTAAATTTTTTAGGTTAAATTTTCTTCCTGCTGATATATATAGGAATTTTACTACTTTAT	2980
Db	1960	CTTTTAAATTTTTTAGGTTAAATTTTCTTCCTGCTGATATATATAGGAATTTTACTACTTTAT	2019
Qy	2981	GTCCTGCTCTTAAACTACATCTCGAACTCGAGCTCGAGGTATATAACAACAGAGCAC	3040
Db	2020	GTCCTGCTCTTAAACTACATCTCGAACTCGAGCTCGAGGTATATAACAACAGAGCAC	2079
Qy	3041	TTTTTTAGGGCAATTTGAAAACCAACCTACACTCTTCGGTGCTTAGAGAGATCTCGTGCT	3100
Db	2080	TTTTTTAGGGCAATTTGAAAACCAACCTACACTCTTCGGTGCTTAGAGAGATCTCGTGCT	2139
Qy	3101	CCCCAAATTAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAATGATTGCTTTCTTT	3160
Db	2140	CCCCAAATTAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAATGATTGCTTTCTTT	2199
Qy	3161	TCCTGGTGATATCTGTGCTTCTCATTAATTCTGAAAGCTGCAATATTTTAGTAATACCTTC	3220
Db	2200	TCCTGGTGATATCTGTGCTTCTCATTAATTCTGAAAGCTGCAATATTTTAGTAATACCTTC	2259
Qy	3221	GGGATCACTGTCCCCCATCTTCGGTGTAGAGCAAAGTGAAGAGTTTAAAGGAGGAAGAA	3280
Db	2260	GGGATCACTGTCCCCCATCTTCGGTGTAGAGCAAAGTGAAGAGTTTAAAGGAGGAAGAA	2319
Qy	3281	GAAGAAGAACTGTCTTACACCACCTTGAGCTCAGACCTCTAAAAACCTGTATTTCCCTTATGAT	3340
Db	2320	GAAGAAGAACTGTCTTACACCACCTTGAGCTCAGACCTCTAAAAACCTGTATTTCCCTTATGAT	2379
Qy	3341	GTCCCTTTTTGAGACACTAAATTTTTTAAATPACTTACTAGCTCTGAAATATATTGATTTTT	3400
Db	2380	GTCCCTTTTTGAGACACTAAATTTTTTAAATPACTTACTAGCTCTGAAATATATTGATTTTT	2439
Qy	3401	ATCACAGTATCTCAGGGTGAATTTAAACCAACTATAGGCCTTTTTCTTGGGATGATTTT	3460
Db	2440	ATCACAGTATCTCAGGGTGAATTTAAACCAACTATAGGCCTTTTTCTTGGGATGATTTT	2499
Qy	3461	CTAGTCTTAAGGTTTGGGGACATTTAAACCTTCAGTACATTTCTGCTFACACAGTTGCATAT	3520
Db	2500	CTAGTCTTAAGGTTTGGGGACATTTAAACCTTCAGTACATTTCTGCTFACACAGTTGCATAT	2559
Qy	3521	TCCAAATTTGATGATGGGAGGAGAGGTGCTTAAAGCTGTAGGCTTTTTCTTTGTACTGC	3580
Db	2560	TCCAAATTTGATGATGGGAGGAGAGAGGTGCTTAAAGCTGTAGGCTTTTTCTTTGTACTGC	2619
Qy	3581	ATTATTAGCATTTAGCTTTAAATATTTTTTAGAGATGTAAACATTTCTCGTTCCTTAGTFC	3640
Db	2620	ATTATTAGCATTTAGCTTTAAATATTTTTTAGAGATGTAAACATTTCTCGTTCCTTAGTFC	2679
Qy	3641	TTACCTAGTCTGAAACATTTTTTAAATAAGAGTTTTTAAATTAAGATTTTGAAAAA AAAA	3700
Db	2680	TTACCTAGTCTGAAACATTTTTTAAATAAGAGTTTTTAAATTAAGATTTTGAAAAA AAAA	2739
Qy	3701	AAA 3703	
Db	2740	AAA 2742	

RESULT 7	BD135177	BD135177	2783 bp	DNA	linear	PAT 18-SEP-2002
LOCUS	Human nucleic acid sequence originating in normal mammary tissue.					
DEFINITION	BD135177					
ACCESSION	BD135177.1	GI:23230122				
VERSION	JP 2002506639-A/24					
KEYWORDS	Homo sapiens (human)					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 2783)					
AUTHORS	Speft,T.; Hintzman,B., Shcmitt,A., Pirarski,C., Duhl,E. and Rosenthal,A.					
TITLE	Human nucleic acid sequence originating in normal mammary tissue					
JOURNAL	Patent: JP 2002506639-A 24 05-MAR-2002;					
COMMENT	METAGEN GESSELLSCHAFT FUER GENOME FORSCHUNG MBH					
	OS Homo sapiens (human)					
	PN JP 2002506639-A/24					
	PD 05-MAR-2002					
	PF 19-MAR-1999 JP 2000536838					
	PR 20-MAR-1998 DE 198 13 835.0					
	PI THOMAS SPEFT,BERND HINTZMAN,ARMIN SCHMITT,CHRISTIAN PIRARSKI,					
	PI EDGAR DUHL,					
	PI ANDRE ROSENTHAL					
	PC C12N15/09,A61K18/00,A61P35/00,A61P43/00,A61P43/00,C07K14/47,					
	PC C07K16/18,					
	PC C12N1/21,C12N5/10,C12P21/02,C12Q1/68,G01N33/68/A61K38/00, PC					
	C12N15/00,					
	PC C12N5/00,A61K37/02					
	CC Human nucleic acid sequence originating in normal mammary CC					
	tissue					
PH Key	Location/Qualifiers					
FT source	1..2783					
FT	/organism='Homo sapiens (human)'					
FEATURES	Location/Qualifiers					
source	1..2783					
	/organism="Homo sapiens"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:9606"					
BASE COUNT	901 a 526 c 609 g 747 t					
ORIGIN						
Query Match	72.8%;	Score 2698.2;	DB 6;	Length 2783;		
Best Local Similarity	99.9%;	Pred. No. 0;				
Matches 2700;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;		
QY 1001	AATGAGCTGAAGCAGCTGGTGGCGAAATCAAATTCATAAATCGACAAAGAGAGAT 1060					
Db 40	AATGAGCTGAAGCAGCTGGTGGCGAAATCAAATTCATAAATCGACAAAGAGAGAT 99					
QY 1061	GTGCCCCCAGGTCCTGAGCTCTGCATCACCATCAGGAAGGGAAGAAATTTCTCCAAT 1120					
Db 100	GTGCCCCCAGGTCCTGAGCTCTGCATCACCATCAGGAAGGGAAGAAATTTCTCCAAT 159					
QY 1121	GAGATAGCCTGGCAGTCGGTTCCACCCCTGCCGAAGATGACTCCCGTGACTCCAGGTT 1180					
Db 160	GAGATAGCCTGGCAGTCGGTTCCACCCCTGCCGAAGATGACTCCCGTGACTCCAGGTT 219					
QY 1181	RAGAGTGAGGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCAGATTCCAGAGCCTCC 1240					
Db 220	RAGAGTGAGGTTCAACAGCCTGTCCATCCCAAGCCACTAAGTCAGATTCCAGAGCCTCC 279					
QY 1241	AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAGAGAGTTTCAGGACCTTGCAGAGAG 1300					
Db 280	AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAGAGAGTTTCAGGACCTTGCAGAGAG 339					
QY 1301	ACCTGGTGGATGTGAGAAGACAGTCTATCCAATGGAGCGCTCTCTTGGCCAAACAGCAG 1360					
Db 340	ACCTGGTGGATGTGAGAAGACAGTCTATCCAATGGAGCGCTCTCTTGGCCAAACAGCAG 399					
QY 1361	GTGTTTTCATCAGCTGCTTCGTTGCTCTATTGCAACAAACAACTCAGTCTTAGGAACA 1420					

Db 400 GTGTTTCACATCAGCTGCTCCGTTGCTCTATTGCAACAACAAACTCAGTCTAGGAACA 459
Qy 1421 TATGCATCTTTACATGGAGAAATCTATTGTGAAGCCTCACITCAATCAACTCTTTAAATCT 1480
Db 460 TATGCATCTTTACATGGAGAAATCTATTGTGAAGCCTCACITCAATCAACTCTTTAAATCT 519
Qy 1481 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACACAAGGATCTATGGCGAAGCAAA 1540
Db 520 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACACAAGGATCTATGGCGAAGCAAA 579
Qy 1541 AATGAAACGAGAGATTTTGGAGAGACCAGCCAGCTTGCMAATGCAAGGGAGACCCCT 1600
Db 580 AATGAAACGAGAGATTTTGGAGAGACCAGCCAGCTTGCMAATGCAAGGGAGACCCCT 639
Qy 1601 CACAGCCAGGGGTAGAAGATGCCCCCTATTGCTTAAGTGGGTGCTCGCTGCGMAAGTATG 1660
Db 640 CACAGCCAGGGGTAGAAGATGCCCCCTATTGCTTAAGGCGGTGCTCGCTGCAAGTATG 699
Qy 1661 GAAGCCAAAGGCTCCTCTCACAGAGAGAAAGACAAAGCCAGCTGAAACCAAGAACTG 1720
Db 700 GAAGCCAAAGGCTCCTCTCACAGAGAGAAAGACAAAGCCAGCTGAAACCAAGAACTG 759
Qy 1721 AGGATCGCTCGGCCACCCCCCACTGAACTTGGAAAGTTCAAGAAAGTGCCTTGGAGAAAGG 1780
Db 760 AGGATCGCTCGGCCACCCCCCACTGAACTTGGAAAGTTCAAGAAAGTGCCTTGGAGAAAGG 819
Qy 1781 ATCAAAATGTCAAAGCCCAATGCGCTCCTGAAAGCAAGAAATCAGCAAGCCCGAAGTTCT 1840
Db 820 ATCAAAATGTCAAAGCCCAATGCGCTCCTGAAAGCAAGAAATCAGCAAGCCCGAAGTTCT 879
Qy 1841 GAGGATGCGATCTAGATCTCAAGAAGCTAAGACGATCTTCTTCACTCAAGAAAGGAAAGC 1900
Db 880 GAGGATGCGATCTAGATCTCAAGAAGCTAAGACGATCTTCTTCACTCAAGAAAGGAAAGC 939
Qy 1901 CGCCCACTCACTGTAGCAGCTTCAATTTCAAAGCACCTCTGTCAAGAGCCCAAAAACCTGTG 1960
Db 940 CGCCCACTCACTGTAGCAGCTTCAATTTCAAAGCACCTCTGTCAAGAGCCCAAAAACCTGTG 999
Qy 1961 TCCCCACCTATCAGAAAGGCTGGAGCATGTCAAGAGAGAGTGAAGAGTGTGTGGGTGGA 2020
Db 1000 TCCCCACCTATCAGAAAGGCTGGAGCATGTCAAGAGAGAGTGAAGAGTGTGTGGGTGGA 1059
Qy 2021 AGAGTTGAGAAAGGAAACAAAGTGGAAATGCCAAGCTTCTTAAGAGAAATGGAAATGTG 2080
Db 1060 AGAGTTGAGAAAGGAAACAAAGTGGAAATGCCAAGCTTCTTAAGAGAAATGGAAATGTG 1119
Qy 2081 GGAAAAACAACCTGGCAAAACAAGAAATCTTAAGAGAGACAGGGAAGAGAAATGAAGAA 2140
Db 1120 GGAAAAACAACCTGGCAAAACAAGAAATCTTAAGAGAGACAGGGAAGAGAAATGAAGAA 1179
Qy 2141 GGTCAATGTTTGGAGATGGAGAAATGAGAAATCTTGTAGAAAAATGGTGCAGACTCCGATGAA 2200
Db 1180 GGTCAATGTTTGGAGATGGAGAAATGAGAAATCTTGTAGAAAAATGGTGCAGACTCCGATGAA 1239
Qy 2201 GATGATAACAGCTTCTTCAAAACACAACTCCCAAGAAATGGTGCAGACTCCGATGAA 2260
Db 1240 GATGATAACAGCTTCTTCAAAACACAACTCCCAAGAAATGGTGCAGACTCCGATGAA 1299
Qy 2261 AGTTTGTAGACAACACCTTTGCTGAAGAATTCATCTACTCAGAAATCAGAAATCCAGGAT 2320
Db 1300 AGTTTGTAGACAACACCTTTGCTGAAGAATTCATCTACTCAGAAATCAGAAATCCAGGAT 1359
Qy 2321 GTGGAACCTCTGGGAGGGAAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 2380
Db 1360 GTGGAACCTCTGGGAGGGAAGAGTGGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGA 1419
Qy 2381 AATCGGATTTATGATGAGGATGAGGATGAAGAGTGAACAATTTGCAATGTATGCTGGGCCCT 2440
Db 1420 AATCGGATTTATGATGAGGATGAGGATGAAGAGTGAACAATTTGCAATGTATGCTGGGCCCT 1479
Qy 2441 AAATTCATGTTAGTGTAGCGAGCCACTGCGCTTTGTCMAAATGTGATGCACATAAGCAG 2500
Db 1480 AAATTCATGTTAGTGTAGCGAGCCACTGCGCTTTGTCMAAATGTGATGCACATAAGCAG 1539

Qy 2501 GTATCCAGCATGAATGTAAATTTACTTTGGAGTAACCTTTGGAAAGAAATTCCTTCTTAA 2560
Db 1540 GTATCCAGCATGAATGTAAATTTACTTTGGAGTAACCTTTGGAAAGAAATTCCTTCTTAA 1599
Qy 2561 AATCAAAAAACAAAAACAAAAACAAAAACACATCTCTAAATACTAGAGATAACTTTAC 2620
Db 1600 AATCAAAAAACAAAAACAAAAACAAAAACACATCTCTAAATACTAGAGATAACTTTAC 1659
Qy 2621 TTAATTTCTTCATTTTAGCAGTGATATGCATGAAGTCTGTAAAGCTTGTAACTGGG 2680
Db 1660 TTAATTTCTTCATTTTAGCAGTGATATGCATGAAGTCTGTAAAGCTTGTAACTGGG 1719
Qy 2681 AAATATTCACCTGATAATAGCCAGATCTACTGTATTCCTCAAAAAGGCAATATTAAAGT 2740
Db 1720 AAATATTCACCTGATAATAGCCAGATCTACTGTATTCCTCAAAAAGGCAATATTAAAGT 1779
Qy 2741 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAAG 2800
Db 1780 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAAG 1839
Qy 2801 GAATTTAGGGGCTTAAACATTTACGACTGAATGCACCTTTAGTATAAAGGCGCACAGTTTGT 2860
Db 1840 GAATTTAGGGGCTTAAACATTTACGACTGAATGCACCTTTAGTATAAAGGCGCACAGTTTGT 1899
Qy 2861 TATTTTAAATGAATACCAATTTAATTTTGTAGTATTTTACCTGTTAAAGAGATTAATTTAGT 2920
Db 1900 TATTTTAAATGAATACCAATTTAATTTTGTAGTATTTTACCTGTTAAAGAGATTAATTTAGT 1959
Qy 2921 CTTTAAATTTTGTAGTTAATTTTCTGCTGTGATATATATAGGAAATTTACTACTTTAT 2980
Db 1960 CTTTAAATTTTGTAGTTAATTTTCTGCTGTGATATATATAGGAAATTTACTACTTTAT 2019
Qy 2981 GTCCTGCTCTTAACTACATCCTGAATCGAGCTCCTGAGGTATAATACAACAGAGCAC 3040
Db 2020 GTCCTGCTCTTAACTACATCCTGAATCGAGCTCCTGAGGTATAATACAACAGAGCAC 2079
Qy 3041 TTTTGAAGCAATTTGAAACCAACCTACACTCTTCGGTGTCTTAGAGAGATCTGCTGTCT 3100
Db 2080 TTTTGAAGCAATTTGAAACCAACCTACACTCTTCGGTGTCTTAGAGAGATCTGCTGTCT 2139
Qy 3101 CCCAAATAAGCTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTCCTTTCTTT 3160
Db 2140 CCCAAATAAGCTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTCCTTTCTTT 2199
Qy 3161 TCTGTGTATCTGCTCTCATTAATTAAGAGCTGCAATATTTTAGTAATACCTTC 3220
Db 2200 TCTGTGTATCTGCTCTCATTAATTAAGAGCTGCAATATTTTAGTAATACCTTC 2259
Qy 3221 GGGATCACCTGTCCGCCATCTTCCGTGTTAGAGTAAAGTGAAGAGTTTAAAGGAGAGAA 3280
Db 2260 GGGATCACCTGTCCGCCATCTTCCGTGTTAGAGTAAAGTGAAGAGTTTAAAGGAGAGAA 2319
Qy 3281 GAAAGAACTGTCTTACACACTTGAGCTCAGACCTCTTAAACCCCTGTATTTCCCTTATGAT 3340
Db 2320 GAAAGAACTGTCTTACACACTTGAGCTCAGACCTCTTAAACCCCTGTATTTCCCTTATGAT 2379
Qy 3341 GTCCCTTTTTCAGACACTAATTTTTTAAATCTACTAGCTCTGAAATATATTGATTTTT 3400
Db 2380 GTCCCTTTTTCAGACACTAATTTTTTAAATCTACTAGCTCTGAAATATATTGATTTTT 2439
Qy 3401 ATCAGAGTATTTCTAGGTTGAAATTTAAACCAACTATAGGCCCTTTTCTTGGGATCATTTT 3460
Db 2440 ATCAGAGTATTTCTAGGTTGAAATTTAAACCAACTATAGGCCCTTTTCTTGGGATCATTTT 2499
Qy 3461 CTAGTCTTAAAGTTTGGGACATTTAAACCTTTGAGTACATTTTGTGTACACAGTTGATAT 3520
Db 2500 CTAGTCTTAAAGTTTGGGACATTTAAACCTTTGAGTACATTTTGTGTACACAGTTGATAT 2559
Qy 3521 TCCAAATTTGTATGGATGGAGGGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTTGTACTGCG 3580
Db 2560 TCCAAATTTGTATGGATGGAGGGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTTGTACTGCG 2619

QY 3581 ATTATAGAGATTAGCTTTAAATATTTTTTAGAGATGTAACACATCTGCTTTCTTAGTC 3640
|||||
Db
2620 ATTTATAGAGATTAGCTTTAAATATTTTTTAGAGATGTAACACATCTGCTTTCTTAGTC 2679
|||||
QY 3641 TTACTAGTCTGAACATTTTTTATTCATAAAGATTTTAAATTAATAATTTGAAAAA 3700
|||||
Db 2680 TTACTAGTCTGAACATTTTTTATTCATAAAGATTTTAAATTAATAATTTGAAAAA 2739
|||||
QY 3701 AAA 3703
|||
Db 2740 AAA 2742
|||

RESULT 8
AF157325 3348 bp mRNA linear PRI 04-MAY-2000
LOCUS Homo sapiens sterol regulatory element binding protein 3 (SREBP3)
DEFINITION mRNA, complete cds.
ACCESSION AF157325
VERSION AF157325.1 GI:7688700
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3348)
Huang,Q., Gu,J., Ren,S., Jin,W., Gu,Y., Gu,W., Dong,H., Yu,Y.,
Fu,G., Wang,Y., Chen,Z. and Han,Z.
A novel gene expressed in the human hypothalamus
Unpublished
TITLE 2 (bases 1 to 3348)
JOURNAL Ren,S., Shi,J., Huang,C., Jiang,C., Li,Y., Zhou,J., Yu,Y., Xu,S.,
Wang,Y., Fu,Q., Chen,Z. and Han,Z.
AUTHORS
REFERENCE Direct Submission
Submitted (09-JUN-1999) Chinese National Human Genome Center at
Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
Shanghai 201203, P. R. China
JOURNAL
FEATURES Location/Qualifiers
source
1..3348
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="hypothalamus"
gene 1..3348
/gene="SREBP3"
CDS 691..2064
/gene="SREBP3"
/codon_start=1
/product="sterol regulatory element binding protein 3"
/protein_id="AA67491.1"
/db_xref="GI:7688701"
/translation="MEQKENVPPPEVCITHQEGEKISANENSIAVRSTPAEDDSRDS
QVKEVQPVHPKPLSPDRASSLSSESPPKMKKFOAPARETCVCOKTVVPMERLL
ANQOVFHSICFRCSYCNKLSLCTYASLHGRIVCKPHNQLPKSKNGYDEGFGRPHK
DLWASKNEIEILERPAQLANARETPHQGVDEDAPIAKVGLVAASWEAKASQQEKEN
RPAETKLRITAWPPFTELSSGSALBEGIMSKPKWPEDEISKEPVEDVDLDKKL
RRSSLSKERSPTVAASQSTSVSPKTPVPPPIRKGSMSQSESVGVRVAERKQV
ENAKSKNGNVKQTTWQNKESKGTCKSGKESLEMENILENANGVSDDEDDNSFL
KQSPQEPKSLNWSFVDNFTAFEEFTQNKQSDVELWEGEVVKELSVKEQIKRNY
DEDEEE"
BASE COUNT 1052 a 620 c 703 g 973 t
ORIGIN
Query Match 72.3%; Score 2680; DB 9; Length 3348;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2694; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1001 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAT 1060
|||||
Db 649 AATGAGCTGAAGCCAGTGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAT 708
|||||
QY 1061 GTGCCCCAGGTCTGAGGTCTGCATCACCCTCAGAGGGGAAAGATTTCTGCAAT 1120
|||||

709 GTGCCCCAGGTCCTGAGGTCTGCATCACCCTCAGGAAGGGGAAAGATTTCTGCAAT 768
|||||
QY 1121 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCGGAAGATGACTCCCGTGACTCCGAGTT 1180
|||||
Db 769 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCGGAAGATGACTCCCGTGACTCCGAGTT 828
|||||
QY 1181 AAGAGTGAGGTTCAACAGAGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCCTCC 1240
|||||
Db 829 AAGAGTGAGGTTCAACAGAGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGCCCTCC 888
|||||
QY 1241 AGTCTTTTGAAGTTCTCTCCCAAAGCAATGAAGAAGTTTTCAGGCACCTGCAAGAGAG 1300
|||||
Db 889 AGTCTTTTGAAGTTCTCTCCCAAAGCAATGAAGAAGTTTTCAGGCACCTGCAAGAGAG 948
|||||
QY 1301 ACTGCGTGGAAATGTCAGAAAGACAGTCTATCAATGGAGCGTCTCTTGGCCCAACAGCAG 1360
|||||
Db 949 ACTGCGTGGAAATGTCAGAAAGACAGTCTATCAATGGAGCGTCTCTTGGCCCAACAGCAG 1008
|||||
QY 1361 GTGTTTCACATCAGCTGCTTCGTTGCTCTTATGCAACAACAAACTCAGTCTAGGAACA 1420
|||||
Db 1009 GTGTTTCACATCAGCTGCTTCGTTGCTCTTATGCAACAACAAACTCAGTCTAGGAACA 1068
|||||
QY 1421 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCACTTCAATCAATCTTTTAAATCT 1480
|||||
Db 1069 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCACTTCAATCAATCTTTTAAATCT 1128
|||||
QY 1481 AAGGCAACTATGATGAAGGCTTTGGGCACACACACAGGATCTATGGGCAAGCAAA 1540
|||||
Db 1129 AAGGCAACTATGATGAAGGCTTTGGGCACACACACAGGATCTATGGGCAAGCAAA 1188
|||||
QY 1541 AATGAAAACGAAGAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGAGAGACCCCT 1600
|||||
Db 1189 AATGAAAACGAAGAGATTTTGGAGAGACAGCCAGCTTGCAAAATGCAAGGAGAGACCCCT 1248
|||||
QY 1601 CACAGCCAGGGGTAGAAGATGCCCTATTGCTAAGTGGGTGCTCTGCTGCTCAAGTATG 1660
|||||
Db 1249 CACAGCCAGGGGTAGAAGATGCCCTATTGCTAAGTGGGTGCTCTGCTGCTCAAGTATG 1308
|||||
QY 1661 GAAGCCAGGCGCTCTCTCAGCAGGAGAGGAGACACAGCCAGCTGAAACCAAGAGGCTG 1720
|||||
Db 1309 GAAGCCAGGCGCTCTCTCAGCAGGAGAGGAGACACAGCCAGCTGAAACCAAGAGGCTG 1368
|||||
QY 1721 AGGATCGCTGGCCACCCCCACCTGAACTTGAAGTTCAGGAAGTGCCTTGGAGGAAGG 1780
|||||
Db 1369 AGGATCGCTGGCCACCCCCACCTGAACTTGAAGTTCAGGAAGTGCCTTGGAGGAAGG 1428
|||||
QY 1781 ATCAAAATGTCAAGCCCAAAATGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGATTCCT 1840
|||||
Db 1429 ATCAAAATGTCAAGCCCAAAATGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGATTCCT 1488
|||||
QY 1841 GAGGATGTCGATCTAGATCTGAAGAGGCTTAAGACGATCTTCTCACTGAAGAAAGAGC 1900
|||||
Db 1489 GAGGATGTCGATCTAGATCTGAAGAGGCTTAAGACGATCTTCTCACTGAAGAAAGAGC 1548
|||||
QY 1901 CCCCCATTCACTGTAGCAGCTTCATTTCAAAGCAGCTCTGTCAAGAGCCCAAAACTGTG 1960
|||||
Db 1549 CCCCCATTCACTGTAGCAGCTTCATTTCAAAGCAGCTCTGTCAAGAGCCCAAAACTGTG 1608
|||||
QY 1961 TCCCCACCTATCAGGAAAGGCTGGAGCATGTGTCAGAGCAGAGTGAAGAGTCTGTGGTGGG 2020
|||||
Db 1609 TCCCCACCTATCAGGAAAGGCTGGAGCATGTGTCAGAGCAGAGTGAAGAGTCTGTGGTGGG 1668
|||||
QY 2021 AGAGTTGCAGAAAGGAAACAAGTGGAAAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTG 2080
|||||
Db 1669 AGAGTTGCAGAAAGGAAACAAGTGGAAAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTG 1728
|||||
QY 2081 GAAAAACCAACTGGCAAAACAAAGAACTCTAAAGGAGAGACAGGGAGAGAGTAAGGAA 2140
|||||
Db 1729 GAAAAACCAACTGGCAAAACAAAGAACTCTAAAGGAGAGACAGGGAGAGAGTAAGGAA 1788
|||||
QY 2141 GGTCAATAGTTTGGAGATGGAGAAATGAGAAATCTTGTAGAAAAATGGTGCAGACTCCGATGAA 2200
|||||
Db 1789 GGTCAATAGTTTGGAGATGGAGAAATGAGAAATCTTGTAGAAAAATGGTGCAGACTCCGATGAA 1848
|||||

```
2201 GATGATAACAGCTTCCTCAAAACAAATCTCCACAAGAACCCAAAGTCTCTGAATTGGTCG 2260
Db |
1849 GATGATACAGCTTCCTCAAAACAAATCTCCACAAGAACCCAAAGTCTCTGAATTGGTCG 1908
Qy |
2261 AGTTTGTAGACAACACCTTTTCTGTAAGAATTCTACTCTCAGAAATCGAATAATCCAGGAT 2320
Db |
1909 AGTTTGTAGACAACACCTTTTCTGTAAGAATTCTACTCTCAGAAATCGAATAATCCAGGAT 1968
Qy |
2321 GTGGAACTCTGGGAGGGAAGTGTCAAGAGCTCTCTGTGGAAGAACACAGATAAAGAGA 2380
Db |
1969 GTGGAACTCTGGGAGGGAAGTGTCAAGAGCTCTCTGTGGAAGAACACAGATAAAGAGA 2028
Qy |
2381 AATCGGTATTATGATGAGGATGAGGATCAAGAGTCAAGATGACAAATGCAATGATGCTGGGCTT 2440
Db |
2029 AATCGGTATTATGATGAGGATGAGGATGAAGAGTACAAATGCAATGATGCTGGGCTT 2088
Qy |
2441 AAATTCATCTGTAGTGTAGCGAGCCACATGCGCCCTTTGTCAAAATGTGATGCACATAAGCAG 2500
Db |
2089 AAATTCATCTGTAGTGTAGCGAGCCACATGCGCCCTTTGTCAAAATGTGATGCACATAAGCAG 2148
Qy |
2501 GTATCCCAGCATGAAATGTAATTTACTTGGAGTAACCTTTTGGAAAGAAATTCCTTCTTAA 2560
Db |
2149 GTATCCCAGCATGAAATGTAATTTACTTGGAGTAACCTTTTGGAAAGAAATTCCTTCTTAA 2208
Qy |
2561 AATCAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAA 2620
Db |
2209 AATCAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAA 2268
Qy |
2621 TTAATTTCTTCATTTTAGCAGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2680
Db |
2269 TTAATTTCTTCATTTTAGCAGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2328
Qy |
2691 AAATATCCACCTGATAATAGCCAGATTTCTACTGTATTTCCAAAAGGCAATATTAAGGT 2740
Db |
2329 AAATATCCACCTGATAATAGCCAGATTTCTACTGTATTTCCAAAAGGCAATATTAAGGT 2388
Qy |
2741 AGATAGATGATTAGTAGTATATTTGTTACACACTATTTTGGAAATTAGAAGACATACAGAG 2800
Db |
2389 AGATAGATGATTAGTAGTATATTTGTTACACACTATTTTGGAAATTAGAAGACATACAGAG 2448
Qy |
2801 GAATTTAGGGCTTAAACATACAGCTCAATGCACTTTAGTATATAAAGGGCACATTTGTA 2860
Db |
2449 GAATTTAGGGCTTAAACATACAGCTCAATGCACTTTAGTATATAAAGGGCACATTTGTA 2508
Qy |
2861 TATTTTAAATGAATACCAATTTAATTTTTTAGTATTTACCTGTTAAAGAGATTTATTAGT 2920
Db |
2509 TATTTTAAATGAATACCAATTTAATTTTTTAGTATTTACCTGTTAAAGAGATTTATTAGT 2968
Qy |
2921 CTTTAAATTTTTTAGGTTAATTTCTGCTGTATATATGAGGAATTTACTACTTTAT 2980
Db |
2569 CTTTAAATTTTTTAGGTTAATTTCTGCTGTATATATGAGGAATTTACTACTTTAT 2628
Qy |
2981 GTCCTGCTCTTAACATACATCCTGNACTGACGCTCCTGAGGTATATAACACAGAGCAC 3040
Db |
2629 GTCCTGCTCTTAACATACATCCTGNACTGACGCTCCTGAGGTATATAACACAGAGCAC 2688
Qy |
3041 TTTTGTAGGCAATTTGAAACAAACCACTACCTTTTCGGTGTCTTAGAGAGATCTGCTCTCT 3100
Db |
2689 TTTTGTAGGCAATTTGAAACAAACCACTACCTTTTCGGTGTCTTAGAGAGATCTGCTCTCT 2748
Qy |
3101 CCCAAATTAAGCTTTGTATCTGCCAGTGAATTTACTGTACTTCCAAATGATTGCTTTCTTT 3160
Db |
2749 CCCAAATTAAGCTTTGTATCTGCCAGTGAATTTACTGTACTTCCAAATGATTGCTTTCTTT 2808
Qy |
3161 TCTGGT-GATATCTGCTCTCTCATTAATTACTGAAAGCTGCAATATTTTGTAGTAATACCTT 3219
Db |
2809 TCTGGTGATATCTGCTCTCTCATTAATTACTGAAAGCTGCAATATTTTGTAGTAATACCTT 2868
Qy |
3220 CGGGATCACTGCTCCCATCTCTTCGGTGTAGAGCAAAAGTGAAGGTTTAAAGGGAGAGA 3279
Db |
2869 CGGGATCACTGCTCCCATCTCTTCGGTGTAGAGCAAAAGTGAAGGTTTAAAGGGAGAGA 2928
```

```
Qy 3280 AGAAGAAGCTGTCTTACACACTTGAGCTCAGACCTCTAAACCCCTGATTTCCTTATGA 3339
Db |
2929 AGAAGAAGCTGTCTTACACACTTGAGCTCAGACCTCTAAACCCCTGATTTCCTTATGA 2988
Qy 3340 TGTCCTCTTTTGGAGACACTAATTTTAAATACTTACTAGCTCTGAAATATATTGATTTT 3399
Db |
2989 TGTCCTCTTTTGGAGACACTAATTTTAAATACTTACTAGCTCTGAAATATATTGATTTT 3048
Qy 3400 TATCAGATATTCTCAGGTGAAATTAACCAACTATAGCCCTTTTCTTGGGATGATTT 3459
Db |
3049 TATCAGATATTCTCAGGTGAAATTAACCAACTATAGCCCTTTTCTTGGGATGATTT 3108
Qy 3460 TCTAGTCTTAAAGTTTGGGACATTATAACCTGAGTACATTGTTGTACACAGTTGATA 3519
Db |
3109 TCTAGTCTTAAAGTTTGGGACATTATAACCTGAGTACATTGTTGTACACAGTTGATA 3168
Qy 3520 TTCCAAATTTGATGGATGGAGGAGGTCTTAAAGCTGTAGGCTTTCTTTTGTACTG 3579
Db |
3169 TTCCAAATTTGATGGATGGAGGAGGTCTTAAAGCTGTAGGCTTTCTTTTGTACTG 3228
Qy 3580 CATTATAGAGATTAGCTTTAATATTTTATAGAGATGTAAACATCTCTGCTTTCTTAGT 3639
Db |
3229 CATTATAGAGATTAGCTTTAATATTTTATAGAGATGTAAACATCTCTGCTTTCTTAGT 3288
Qy 3640 CTTACCTAGTCTGAACATTTTATTCATAAGATTTTTAATTAATAATTTGAAAAAATAA 3699
Db |
3289 CTTACCTAGTCTGAACATTTTATTCATAAGATTTTTAATTAATAATTTGAAAAAATAA 3348

RESULT 9
AK000057 2667 bp mRNA linear PRI 22-FEB-2000
LOCUS Homo sapiens cDNA FLJ20050 fis, clone COL00688.
DEFINITION AK000057
ACCESSION AK000057
VERSION AK000057.1 GI:7019894
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2667)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u.tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1..2667
location=Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cdname="COL00688"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"
10..1380
/note="unnamed protein product"
/codon_start=1

CDS
```

/protein_id="BAA90914.1"									
/db_xref="GI:7019895"									
/translation="MEQKENVPPGPEVCIITHOGEKISANENSILAVRSTPAEDSRDS									
QYKSEVQVPHKPLSPDSSRASSLESSPPKANKFOAPARETCVEQKTVYEMERLLA									
NQOVPHISCFRCSYCNKKUSLGYALSHGRYYCKPFPNQLFKSNKYNDGFGHRPHKG									
LWASKNENEEIIRPQLANARETSPGVDEPIAKVGLAASMEAKASSQEQEKEDK									
PAETKKLRIAMPPTPELSSGSALEBEGIKMSKPKWPPDEDEISKPEVPEDVDLKKLR									
RSSLSKERSPFTVAASFSTSVKSPKTVSPPIRKWMSSEOSEESVGGRVAERKQVE									
NAKASKNGNVGKTTWONKESKGETCKRSEKSHSLEMENENLVENGADSDDEDSFLK									
QQSPQPKSLUNSSFVDNITFABEFTTQNKQSDVELWEGEVKEUSVEEQIKRNIYD									
EDDEE"									
BASE COUNT	865 a	512 c	570 g	720 t					
ORIGIN									
Query Match	70.3%;	Score	2604.4;	DB 9;	Length	2667;			
Best Local Similarity	99.3%;	Pred. No. 0;							
Matches 2654;	Conservative	0;	Mismatches	1;	Indels	19;	Gaps	3;	
QY	1038	ATAAAATGGAGCAAAAGGAAATGTGCCCCACAGTCTCTGAGGTCTCGATCACCATCAGG	1097						
Db	5	ATAAAATGGAGCAAAAGGAAATGTGCCCCACAGTCTCTGAGGTCTCGATCACCATCAGG	64						
QY	1098	AAGGGGAAAAAGATTTCTGCAAAATGAGAATAGCCTGGCAGTCCGTTCCACCCCTGCCGAAG	1157						
Db	65	AAGGGGAAAAAGATTTCTGCAAAATGAGAATAGCCTGGCAGTCCGTTCCACCCCTGCCGAAG	124						
QY	1158	ATGACTCCCGTGACTCCAGGTTAAGAGTGAAGTTCAACAGCCTGTCCATCCCAAGCCAC	1217						
Db	125	ATGACTCCCGTGACTCCAGGTTAAGAGTGAAGTTCAACAGCCTGTCCATCCCAAGCCAC	184						
QY	1218	TAAGTCCAGATTCACAGGCTCCAGTCTTTCTCAAAAGTCTCTCTCCCAAGCAATCAAGA	1277						
Db	185	TAAGTCCAGATTCACAGGCTCCAGTCTTTCTCAAAAGTCTCTCTCCCAAGCAAT--GA	241						
QY	1278	AGTTTCAGGCACCTGCAAGAGAGACCTGGTGGAAATGTGAGAAGACAGTCTATCCAATGG	1337						
Db	242	AGTTTCAGGCACCTGCAAGAGAGACCTGGTGGAAATGTGAGAAGACAGTCTATCCAATGG	301						
QY	1338	AGCGTCTCTTGGCCAAACAGCAGGTTTTCACATCAGCTGCTTCCGTTGCTCTCTATTGCA	1397						
Db	302	AGCGTCTCTTGGCCAAACAGCAGGTTTTCACATCAGCTGCTTCCGTTGCTCTCTATTGCA	361						
QY	1398	ACAACAAACTCAGTCTAGGAACATATGCATCTTTACATGGAAGAAATCTATTGTAGGCCCTC	1457						
Db	362	ACAACAAACTCAGTCTAGGAACATATGCATCTTTACATGGAAGAAATCTATTGTAGGCCCTC	421						
QY	1458	ACTTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGCTTTGGGCACAGACCAC	1517						
Db	422	ACTTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGCTTTGGGCACAGACCAC	481						
QY	1518	ACAAGGATCTATGGGCAAGCAAAAATGAAACGAAGAGATTTTGGAGAGACCCAGCCAGC	1577						
Db	482	ACAAGGATCTATGGGCAAGCAAAAATGAAACGAAGAGATTTTGGAGAGACCCAGCCAGC	541						
QY	1578	TTGCAATGCAAGGAGAGACCCCTCAGCCAGGGGTAGAAGATGCCCCATTGTCTAAGG	1637						
Db	542	TTGCAATGCAAGGAGAGACCCCTCAGCCAGGGGTAGAAGATGCCCCATTGTCTAAGG	601						
QY	1638	TGGGTGTCTGGCTGCAAGTATGGAAGCAAGCCCTCTCTCAGCAGGAGAGAGAGACA	1697						
Db	602	TGGGTGTCTGGCTGCAAGTATGGAAGCAAGCCCTCTCTCAGCAGGAGAGAGAGACA	661						
QY	1698	AGCCAGCTGAAACCAAGAAGCTGAGGATCGCCTGGCCACCCCCCTGAACCTTGGAAAGTT	1757						
Db	662	AGCCAGCTGAAACCAAGAAGCTGAGGATCGCCTGGCCACCCCCCTGAACCTTGGAAAGTT	721						
QY	1758	CAGGAAGTGCCTTGGAGGAAGGATCAAAATGTCAAAGCCCAATGGCCTCTCTGAAGAGC	1817						
Db	722	CAGGAAGTGCCTTGGAGGAAGGATCAAAATGTCAAAGCCCAATGGCCTCTCTGAAGAGC	781						
QY	1818	AAATCAGCAGCCGAGTCTCTGAGGATGTCGATCTAGATCTGAGAGAGCTTAAGACGAT	1877						
Db	782	AAATCAGCAGCCGAGTCTCTGAGGATGTCGATCTAGATCTGAGAGAGCTTAAGACGAT	841						

1878	CTTCTTCACTGAAGAAAGAGCGCCCATTCAGTGTAGCAGCTTCATTTCAAAGCACCT	1937
842	CTTCTTCACTGAAGAAAGAGCGCCCATTCAGTGTAGCAGCTTCATTTCAAAGCACCT	901
1938	CTGTCAAGAGCCCAAAAACCTGTGTCCCCACCATTCATCAGAAAAGGCTGGAGCATGTCCAGAGC	1997
902	CTGTCAAGAGCCCAAAAACCTGTGTCCCCACCATTCATCAGAAAAGGCTGGAGCATGTCCAGAGC	961
1998	AGAGTGAAGAGTCTGTGGTGGAGAGGTTGAGAAAAGGAAACAAGTGGAAAATGCCAAGG	2057
962	AGAGTGAAGAGTCTGTGGTGGAGAGGTTGAGAAAAGGAAACAAGTGGAAAATGCCAAGG	1021
2058	CTTCTTAAGAAAGATGGGNAATGTGGGAAAAACAACCTGGCAAAAAACAAGAAATCTTAAGAGAG	2117
1022	CTTCTTAAGAAAGATGGGNAATGTGGGAAAAACAACCTGGCAAAAAACAAGAAATCTTAAGAGAG	1081
2118	AGACAGGGAAGAGAGAGTAAAGGAAGTCTAGTCTTGGAGATGGAGAAATGAGAAATCTTGTAG	2177
1082	AGACAGGGAAGAGAGTAAAGGAAGTCTAGTCTTGGAGATGGAGAAATGAGAAATCTTGTAG	1141
2178	AAAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCTCTCAAAACAACAATCTCCACAAG	2237
1142	AAAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCTCTCAAAACAACAATCTCCACAAG	1201
2238	AACCCAGTCTCTGAAATTTGGTTCGAGTTTTTTGTAGACAACACCTTTTGTGAGAAATTCACCTA	2297
1202	AACCCAGTCTCTGAAATTTGGTTCGAGTTTTTTGTAGACAACACCTTTTGTGAGAAATTCACCTA	1261
2298	CTCAGAATCAGAAATCCCAAGATGTGNACTCTGGAGGGAGAAAGTGGTCAAGAGCTCT	2357
1262	CTCAGAATCAGAAATCCCAAGATGTGNACTCTGGAGGGAGAAAGTGGTCAAGAGCTCT	1321
2358	CTGTGGAAGAAACAGATAAAGAGAAATCGGTATATGATGAGGATGAGGATCAAGAGTGCAC	2417
1322	CTGTGGAAGAAACAGATAAAGAGAAATCGGTATATGATGAGGATGAGGATCAAGAGTGCAC	1381
2418	AAATTGCAATGATCGTGGCCCTTAAATTTCAATGTTAGTGTAGCGAGCCACTGCCCTTTGT	2477
1382	AAATTGCAATGATCGTGGCCCTTAAATTTCAATGTTAGTGTAGCGAGCCACTGCCCTTTGT	1441
2478	CAAAATGTGATGCAATTAAGCAGGATATCCAGCATGAAATGTAATTTACTTTGGAAGTAAC	2537
1442	CAAAATGTGATGCAATTAAGCAGGATATCCAGCATGAAATGTAATTTACTTTGGAAGTAAC	1501
2538	TTTGGAAGAAATTCCTTTAAATCAAAAACAACAACAACAACAACAACAACAACAACAATT	2597
1502	TTTGGAAGAAATTCCTTTAAATCAAAAACAACAACAACAACAACAACAACAACAACAATT	1561
2598	CTAAATCTAGAGATAACTTTTACTTTAAATTTCTTCAATTTAGCAGTGATGATGCATAAG	2657
1562	CTAAATCTAGAGATAACTTTTACTTTAAATTTCTTCAATTTAGCAGTGATGATGCATAAG	1621
2658	TGCTGTAAAGGCTTGAATGGGGAAATATTCCACCTGATAATAGCCAGATTTCTACTGTA	2717
1622	TGCTGTAAAGGCTTGAATGGGGAAATATTCCACCTGATAATAGCCAGATTTCTACTGTA	1681
2718	TTCCCAAAAGGCAATATTAAGGTAGATAGATGATTAGTAGTATATTGTTACACACTATTT	2777
1682	TTCCCAAAAGGCAATATTAAGGTAGATAGATGATTAGTAGTATATTGTTACACACTATTT	1741
2778	TGGAATTAGAGAACATACAGAAAGAAATTTAGGGGCTTAAACATTTACGACTGAATGCCACTT	2837
1742	TGGAATTAGAGAACATACAGAAAGAAATTTAGGGGCTTAAACATTTACGACTGAATGCCACTT	1801
2838	TAGTATAAAGGGCACAGTTTGTATATTTTAAATGAATACCAATTTTAAATTTTGTAGTATT	2897
1802	TAGTATAAAGGGCACAGTTTGTATATTTTAAATGAATACCAATTTTAAATTTTGTAGTATT	1861
2898	TACCTGTTAAGAGATTATTTAGTCTTTAAATTTTTTAAATTTTTTGTGCTGTGATAT	2957
1862	TACCTGTTAAGAGATTATTTAGTCTTTAAATTTTTTAAATTTTTTGTGCTGTGATAT	1921

QY	2958	ATATGAGGAATTTACTACTTTATCTCTCTCTAACTACATCTCTGAACTCGACGTCC	3017
Db	1922	ATATGAGGAATTTACTACTTTATCTCTCTCTAACTACATCTCTGAACTCGACGTCC	1981
QY	3018	TCAGGTATAATACACAGAGACACTTTTGGGGCAATTCGAAAGCAACCTACACTCTTCG	3077
Db	1982	TCAGGTATAATACACAGAGACACTTTTGGGGCAATTCGAAAGCAACCTACACTCTTCG	2041
QY	3078	GTGCTTAGAGAGATCTGCTGTCTCCAAATTAAGCTTTTGTATCTGCGAGTAATTTACTG	3137
Db	2042	GTGCTTAGAGAGATCTGCTGTCTCCAAATTAAGCTTTTGTATCTGCGAGTAATTTACTG	2101
QY	3138	TACTCCAAATGATTCCTTTCTGGTGATATCTGCTTCTCATTAATTAAGCAAGC	3197
Db	2102	TACTCCAAATGATTCCTTTCTGGTGATATCTGCTTCTCATTAATTAAGCAAGC	2161
QY	3198	TGCAATATTTTAGTAATACCTTCGGGATCA-----CTGTCCCCCATCTTCGGGTGAGA	3251
Db	2162	TGCAATATTTTAGTAATACCTTCGGGATCACTAGAGCTGTCCCCCATCTTCGGGTG---	2217
QY	3252	GCAAAGTGAAGATTTAAAGGAGGAAGAAAGAACTGCTTTACACCTTTGAGCTCAG	3311
Db	2218	-----TGAAAGATTTAAAGGAGGAAGAAAGAACTGCTTTACACCTTTGAGCTCAG	2271
QY	3312	ACCTCTAAACCTGTATTTCCCTTATGATGTCCCTTTTGGAGACACTAATTTTAAATA	3371
Db	2272	ACCTCTAAACCTGTATTTCCCTTATGATGTCCCTTTTGGAGACACTAATTTTAAATA	2331
QY	3372	CTTACTAGCTCTGAAATATATTTATTCACAGATTTCTCAGGTTGAAATTTAAACCA	3431
Db	2332	CTTACTAGCTCTGAAATATATTTATTCACAGATTTCTCAGGTTGAAATTTAAACCA	2391
QY	3432	ACTATAGCCCTTTTCTGGGATGATTTCTAGTCTTAAAGTTTGGGACATTATAACT	3491
Db	2392	ACTATAGCCCTTTTCTGGGATGATTTCTAGTCTTAAAGTTTGGGACATTATAACT	2451
QY	3492	TGAGTACATTTGTTGTACACAGTTGATTTCCAAATTTGATGGATGGAGGAGAGTGT	3551
Db	2452	TGAGTACATTTGTTGTACACAGTTGATTTCCAAATTTGATGGATGGAGGAGAGTGT	2511
QY	3552	CTTAAGCTGTAGGCTTTCTTTGTACTGCATTTATAGAGATTTAGCTTTAATATTTTAA	3611
Db	2512	CTTAAGCTGTAGGCTTTCTTTGTACTGCATTTATAGAGATTTAGCTTTAATATTTTAA	2571
QY	3612	GAGATGTAAAAACATTCGCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATTCATAAA	3671
Db	2572	GAGATGTAAAAACATTCGCTTTCTTAGTCTTACCTAGTCTGAAACATTTTATTCATAAA	2631
QY	3672	AGATTTTAAATTTGAAAAAATAAAAAA	3705
Db	2632	AGATTTTAAATTTGAAAAAATAAAAAA	2665
RESULT 10	AK000372	LOCUS	
DEFINITION		2536 bp	linear
ACCESSION	AK000372	2536 bp	linear
VERSION	AK000372	2536 bp	linear
KEYWORDS	AK000372	2536 bp	linear
SOURCE	AK000372	2536 bp	linear
ORGANISM	AK000372	2536 bp	linear
REFERENCE	AK000372	2536 bp	linear
AUTHORS	AK000372	2536 bp	linear
TITLE	AK000372	2536 bp	linear
JOURNAL	AK000372	2536 bp	linear
REFERENCE	AK000372	2536 bp	linear
AUTHORS	AK000372	2536 bp	linear

Shibahara, T., Tanaka, T. and Nakamura, Y.	
Direct Submission	
Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,	
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,	
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,	
Tel:81-3-5449-5286, Fax:81-3-5449-5416)	
NEDO human cDNA sequencing project supported by Ministry of	
International Trade and Industry of Japan; cDNA full insert	
sequencing: Research Association for Biotechnology; cDNA library	
construction, 5'- & 3'-end one pass sequencing: Department of	
Virology and Human Genome Center, Institute of Medical Science,	
University of Tokyo (partly supported by Science and Technology	
Agency).	
Location/Qualifiers	
1. .2536	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone="HEP17877"	
/cell_line="HepG2"	
/cell_type="hepatoma"	
/clone_lib="HEP"	
/note="cloning vector pME18SFL3"	
85. .2199	
/note="unnamed protein product"	
/codon_start=1	
/protein_id="BAA91120.1"	
/db_xref="GI:7020416"	
/translation="MESSPPNRRQWTSLSRVTAKELSLVNKNKSSAIVEIPSKYQKA	
ABETINMKRSNTENLSQFRGLTULVKKWENPLGAEHSHTDSRNSSTIEIRRAD	
HPFATVSHAAGAKAQDEQIHPRSLRPPEALVQGRPHIKOGEYKIDRSTESKK	
MENCGESRHEVEKSEISENTDASGIEKYNVPLNRLKMMFEKSPPTQIKLRASRS	
ASGRKISNSVSLDLEITGQOLSSSTIDSEKNESRNLELPLRSETSIKDMAKYQA	
AVSKOSSNTYNTEILKASGGEIKHMKOKENVPPGECVITHOGEKISANESLAV	
RSTPREDRSQVSEVQVQVPHRPSLPDRASSLSSESSPPKMKKKQAPARETCVE	
CKQTYMERLLANQVFIHSCFRPSYCNKLISLGTYSALHGRIVYKPHENGLFKSG	
NYDEGFRPHKDLWAKNEEIEILRPAQLANARETPHSGVEAPAKVGVLAASM	
EAKASQOEKEDKPAETKRLIAWPPTELSSGSGALBEGKMSKPKWPEDEISKPE	
VPEDVDLKLKLRSSLSKRSRPTVAASFQSTSVKSPKTVSPDPRKGSWSEOSE	
SVGGVAVRRKQVENAKSKNGVCKTTWQNKSESGETGKRSKSHSLEMENELVEN	
GADSDSDNSFLKQHLKHNPSL	
BASE COUNT 863 a 551 c 612 g 510 t	
ORIGIN	
Query Match 67.9%; Score 2515.4; DB 9; Length 2536;	
Best Local Similarity 99.7%; Pred. No. 0;	
Matches 2530; Conservative 0; Mismatches 6; Indels 1; Gaps 1;	
QY 53 AGAGCGCCGGGACCTGTGACAGGGCTGTGACAGCGCAGAGGAAAGCGGCTTTTAGCCA 112	
Db 1 AGAGCGCCTTGACCTGTGACAGGGCTGTGACAGCGCAGAGGAAAGCGGCTTTTAGCCA 60	
QY 113 GGTATTTTCAGTCTGTAGACAAGATGGAATCATCTCCATTTAATAGACGGCAATGGACC 172	
Db 61 GGTATTTTCAGTCTGTAGACAAGATGGAATCATCTCCATTTAATAGACGGCAATGGACC 120	
QY 173 TCACATATCATTGAGGGTAACAGCCAAAGAACTTTCTCTGTGTCACAAAGAACAAAGTCATCG 232	
Db 121 TCACATATCATTGAGGGTAACAGCCAAAGAACTTTCTCTGTGTCACAAAGAACAAAGTCATCG 180	
QY 233 GGTATTTGGAATAATTTCTCCAAAGTACCAGAAAGCAGCTGGAAGAACAAACATGCGAAG 292	
Db 181 GGTATTTGGAATAATTTCTCCAAAGTACCAGAAAGCAGCTGGAAGAACAAACATGCGAAG 240	
QY 293 AAGAGAAGTAAACCCGAAAAATCTCTCCAGCAGCTTTAGAAAAGGGGACCTGACTGTGTTA 352	
Db 241 AAGAGAAGTAAACCCGAAAAATCTCTCCAGCAGCTTTAGAAAAGGGGACCTGACTGTGTTA 300	
QY 353 AAGAGAAGTGGGAAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACCGAAC 412	
Db 301 AAGAGAAGTGGGAAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACCGAAC 360	
QY 413 ACCAGCAGCTGAGATTAGGCACAGACACACCATCTCTCTGCTGAAGTGACAGCCAGCT 472	

Db 361 AGCAGCACTGAGATTAGGCACAGCAGACCATCTCTCTGCTGAGTGCACAGCCAGCT 420
QY 473 GCTTCTGGAGCCAAAGCTGACCAAGAGAACAAATCCACCCAGATCTTATAGATCTCAGGTCA 532
Db 421 GCTTCTGGAGCCAAAGCTGACCAAGAGAACAAATCCACCCAGATCTTATAGATCTCAGGTCA 480
QY 533 CTTCTCTGAAGCCCTCGTTGAGGCTGCATATCCACATCAAGGACGGTGAAGATCTTAAA 592
Db 481 CTTCTCTGAAGCCCTCGTTGAGGCTGCATATCCACATCAAGGACGGTGAAGATCTTAAA 540
QY 593 GACCACCTCAACAGAAAGTAAAGAAATGGAAATTTGTCTAGGAGAAATCCAGGATGAAGTA 652
Db 541 GACCACCTCAACAGAAAGTAAAGAAATGGAAATTTGTCTAGGAGAAATCCAGGATGAAGTA 600
QY 653 GAAAAATCAGAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAATATATATGTT 712
Db 601 GAAAAATCAGAAATCAGTGAAGAACACAGATGCTTCGGGCAAAATAGAGAAATATATATGTT 660
QY 713 CCCTGAAACAGGCTTAAGATGATGTTTGAAGAGGTGAACCAACTCAAACTAAGATTTCTC 772
Db 661 CCCTGAAACAGGCTTAAGATGATGTTTGAAGAGGTGAACCAACTCAAACTAAGATTTCTC 720
QY 773 CGGCCCCAAGCCGAAGTGAAGTGAAGGAAGATCTCTGAAACAGCTATTCTCTAGAT 832
Db 721 CGGCCCCAAGCCGAAGTGAAGTGAAGGAAGATCTCTGAAACAGCTATTCTCTAGAT 780
QY 833 GACCTGGAATAGGCCAGGTGAGTTGTCATCTTTACATTTTACATTCGAGAGAAAAATGAG 892
Db 781 GACCTGGAATAGGCCAGGTGAGTTGTCATCTTTACATTTTACATTCGAGAGAAAAATGAG 840
QY 893 AGTAGAGAAATCTGAACTTCACGCCCTCTCAGAAACCTCTATATAAGGATCGAATGGCC 952
Db 841 AGTAGAGAAATCTGAACTTCACGCCCTCTCAGAAACCTCTATATAAGGATCGAATGGCC 900
QY 1012 AAGTACAGGACGCTGTGTCACCAACAAAGCAGCTCAACCAATATACAAATGAGCTGAAA 1012
Db 901 AAGTACAGGACGCTGTGTCACCAACAAAGCAGCTCAACCAATATACAAATGAGCTGAAA 960
QY 1013 GCCAGTGGTGGCAAAATCAAAATTTCAATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGT 1072
Db 961 GCCAGTGGTGGCAAAATCAAAATTTCAATAAATGGAGCAAAAGGAGAAATGTGCCCCAGGT 1020
QY 1073 CTTGAGGCTGTGCATCACCATCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCCTG 1132
Db 1021 CTTGAGGCTGTGCATCACCATCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCCTG 1080
QY 1133 GCAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTGAATCCCGGTTAAGAGTGAAGTT 1192
Db 1081 GCAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTGAATCCCGGTTAAGAGTGAAGTT 1140
QY 1193 CAACAGCCTGTGCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCCTCCAGTCTTTCTGAA 1252
Db 1141 CAACAGCCTGTGCATCCCAAGCCACTAAGTCCAGATTCCAGAGCCCTCCAGTCTTTCTGAA 1200
QY 1253 AGTTCTCTCCCAAGCAATGAAGAAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAA 1312
Db 1201 AGTTCTCTCCCAAGCAATGAAGAAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAA 1260
QY 1313 TGTCAAGACAGTCTATCAATGGAAGGCTCTCTGGCCAAACAGAGGTGTTTCAATC 1372
Db 1261 TGTCAAGACAGTCTATCAATGGAAGGCTCTCTGGCCAAACAGAGGTGTTTCAATC 1320
QY 1373 AGCTGCTTCGGTTCCTTATTCACACCAACTCAGTCTAGGAACATATGCACTTTTA 1432
Db 1321 AGCTGCTTCGGTTCCTTATTCACACCAACTCAGTCTAGGAACATATGCACTTTTA 1380
QY 1433 CATGGAAGAAATCTATTGTAAGCCCTCACTTCAATCAACTCTTTAAATCTAAGGGCACTAT 1492
Db 1381 CATGGAAGAAATCTATTGTAAGCCCTCACTTCAATCAACTCTTTAAATCTAAGGGCACTAT 1440
QY 1493 GATGAAGGCTTTGGGCACAGACACAAAGGATCTATGGGCAAGCAAAATGAAACCGAA 1552

1441 GATGAAGGCTTTGGGCACAGACACAAAGGATCTATGGGCAAGCAAAATGAAACCGAA 1500
1553 GAGATTTTGGAGAGACAGCCAGCTTTGCAAAATGCAAGGGAGAGCCCTCAGAGCCAGGG 1612
1501 GAGATTTTGGAGAGACAGCCAGCTTTGCAAAATGCAAGGGAGAGCCCTCAGAGCCAGGG 1560
1613 GTAGAAGATGCCCTATTGCTTAAGSTGGTGTCTGCTGCTCAAGTATGGAAGCAAGGCC 1672
1561 GTAGAAGATGCCCTATTGCTTAAGSTGGTGTCTGCTGCTCAAGTATGGAAGCAAGGCC 1620
1673 TCCTCTCAGCAGGAGAGAGAAAGCAAGCCAGCTGAAACCAAGAAAGCTGAGGATCGCTGG 1732
1621 TCCTCTCAGCAGGAGAGAGAAAGCAAGCCAGCTGAAACCAAGAAAGCTGAGGATCGCTGG 1680
1733 CCACCCCCCACTGAACTTTGGAAGTTTCAGAAAGTTCCTTGAAGGAGGGATCAAAATGTCA 1792
1681 CCACCCCCCACTGAACTTTGGAAGTTTCAGAAAGTTCCTTGAAGGAGGGATCAAAATGTCA 1740
1793 AAGCCCAATGCCCTCTCTGAAGACGAAATCAGCAAGCCGAGGTTCTTGAGGATGTGAT 1852
1741 AAGCCCAATGCCCTCTCTGAAGACGAAATCAGCAAGCCGAGGTTCTTGAGGATGTGAT 1800
1853 CTAGATCTGAAGAAAGCTTAAGACGATCTTTCTCACTGAAGGAAAGCCGCCCTTCACT 1912
1801 CTAGATCTGAAGAAAGCTTAAGACGATCTTTCTCACTGAAGGAAAGCCGCCCTTCACT 1860
1913 GTAGCAGGCTTCATTTTCAAGACACCTCTGTCAAGGACCCCAAAATCTGTGCCCCCTATC 1972
1861 GTAGCAGGCTTCATTTTCAAGACACCTCTGTCAAGGACCCCAAAATCTGTGTGCCCCCTATC 1920
1973 AGGAAAGGCTCGAGCATGTCTAGACAGAGTGAAGTCTGTGGTGGAGAGTTCGACAA 2032
1921 AGGAAAGGCTCGAGCATGTCTAGACAGAGTGAAGTCTGTGGTGGAGAGTTCGACAA 1980
2033 AGGAAACAAAGTGGAAATGCCCAGGCTTCTAAGAAAGATGGGAATGGGAAACAAACC 2092
1981 AGGAAACAAAGTGGAAATGCCCAGGCTTCTAAGAAAGATGGGAATGGGAAACAAACC 2040
2093 TGGCAAAACAAAGATCTAAGAGGAGACAGGAGAGAGTAAAGGAGGTCTAGTTTG 2152
2041 TGGCAAAACAAAGATCTAAGAGGAGACAGGAGAGAGTAAAGGAGGTCTAGTTTG 2100
2153 GAGATGGAGAAATGGAATCTTTGTAGAAAATGGTCAGACTCCGATGAAGATGATAACAGC 2212
2101 GAGATGGAGAAATGGAATCTTTGTAGAAAATGGTCAGACTCCGATGAAGATGATAACAGC 2160
2213 TTCTCTCAAAACAAATCTCCAAAGAACCCCAAGTCTCTGAAATGGTTCAGTTTGTAGAC 2272
2161 TTCTCTCAAAACAAAC-ATCTCCAAAGAACCCCAAGTCTCTGAAATGGTTCAGTTTGTAGAC 2219
2273 AACACCTTTGCTGAGAAATTCACCTACTCAGAAATCCAGGATGTGGAACCTCTGG 2332
2220 AACACCTTTGCTGAGAAATTCACCTACTCAGAAATCCAGGATGTGGAACCTCTGG 2279
2333 GAGGAGAAATGGTCAAAAGAGCTCTCTGTGGAAGAACAGATGAAAGAGAAATCGGTATTAT 2392
2280 GAGGAGAAATGGTCAAAAGAGCTCTCTGTGGAAGAACAGATGAAAGAGAAATCGGTATTAT 2339
2393 GATGAGATGAGGATGAGAGTGAAGTGAAGTTCGAATGATGTGGCCCTTAAATTCATGTTA 2452
2340 GATGAGATGAGGATGAGAGTGAAGTGAAGTTCGAATGATGTGGCCCTTAAATTCATGTTA 2399
2453 GTGTTAGGAGCCACTGCCCTTTGTCAAAATGTGATGCACATAGCAGGTTATCCAGCAT 2512
2400 GTGTTAGGAGCCACTGCCCTTTGTCAAAATGTGATGCACATAGCAGGTTATCCAGCAT 2459
2513 GAAATGTAATTTACTTGGAACTACTTTGGAAAGAAATTCCTTCTTAATCAAAACAA 2572
2460 GAAATGTAATTTACTTGGAACTACTTTGGAAAGAAATTCCTTCTTAATCAAAACAA 2519
2573 AACAAAAAACACAAAA 2589
2520 AAAAAAAGAAAAA 2536

RESULT 11 AC008147/c	AC008147	115345 bp	DNA	linear	PRI 29-MAR-2003	Direct Submission Submitted (28-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 5 (bases 1 to 115345) Worley,K.C.
LOCUS	AC008147	115345 bp	DNA	linear	PRI 29-MAR-2003	Direct Submission Submitted (29-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Feb 1, 2002 this sequence version replaced gi:15626025. INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu
DEFINITION	AC008147	115345 bp	DNA	linear	PRI 29-MAR-2003	CLONE LENGTH: This sequence does not necessarily represent the entire length of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.
ACCESSION	AC008147.31	GI:18464004				ANNOTATION OF FEATURES: STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts. Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.
VERSION	AC008147.31	GI:18464004				SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.
KEYWORDS	AC008147.31	GI:18464004				QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html .
SOURCE	AC008147.31	GI:18464004				QUALSTAT-REPORT: Location/Qualifiers 1. 115345 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="12" /clone="RP3-405J10" complement(2..2047) /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
ORGANISM	AC008147.31	GI:18464004				FEATURES Source 1. 115345 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="12" /clone="RP3-405J10" complement(2..2047) /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
REFERENCE	AC008147.31	GI:18464004				misc_feature 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
AUTHORS	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
TITLE	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
JOURNAL	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
REFERENCE	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
AUTHORS	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
TITLE	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
JOURNAL	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
REFERENCE	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
AUTHORS	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
TITLE	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
JOURNAL	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
REFERENCE	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"
AUTHORS	AC008147.31	GI:18464004				repeat_region 1. 115345 /note="overlaps bases 1.2020 of clone AC139016" /function="clone overlap" complement(2..79) /rpt_family="AluY" 110..156 /rpt_family="(A)n" 157..470 /rpt_family="AluSg" 471..484 /rpt_family="(A)n" 998..1132 /standard_name="1946S" 2114..2395 /standard_name="133863" 3365..3655 /rpt_family="AluJb"

```
repeat_region 3668..4037
/rpt_family="LTR24"
repeat_region 4041..4067
/rpt_family="(TA)n"
repeat_region complement(4068..4242)
/rpt_family="AluSg/x"
repeat_region complement(4243..4539)
/rpt_family="AluSc"
repeat_region complement(4541..4836)
/rpt_family="AluY"
repeat_region complement(4837..4890)
/rpt_family="AluSg/x"
repeat_region complement(4891..5005)
/rpt_family="AluJo"
repeat_region 5006..5118
/rpt_family="AluSx"
repeat_region 5119..5425
/rpt_family="AluSx"
repeat_region 5426..5585
/rpt_family="AluSx"
repeat_region complement(5586..5690)
/rpt_family="AluJo"
repeat_region 5931..6233
/rpt_family="AluY"
repeat_region 8032..8320
/rpt_family="AluSg"
repeat_region 8632..8934
/rpt_family="AluSg"
repeat_region 9461..9759
/rpt_family="AluSx"
repeat_region 9920..9947
/rpt_family="(TTTGG)n"
repeat_region complement(9948..10229)
/rpt_family="AluSg"
repeat_region 11773..11943
/rpt_family="LTR47A"
repeat_region complement(114835..15109)
/rpt_family="AluSg"
repeat_region complement(15112..15412)
/rpt_family="AluSg"
repeat_region complement(15467..15779)
/rpt_family="AluSp"
repeat_region 15783..15818
/rpt_family="(T)n"
repeat_region 15932..16230
/rpt_family="AluJo"
repeat_region 16339..16637
/rpt_family="AluJb"
repeat_region 16840..16933
/rpt_family="AluY"
repeat_region 16934..16956
/rpt_family="(CAAA)n"
repeat_region 16962..17249
/rpt_family="AluSg"
repeat_region 17251..17288
/rpt_family="AT-rich"
repeat_region 18079..18112
/rpt_family="(TG)n"

Query Match 51.7%; Score 2284.4; DB 9; Length 115345;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2294; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1383 GTTGCTCCTATTGCAACAACAACTCAGTCAGGAACATATGCATCTTTACATGGAAGAA 1442
|||
Db 3197 GTTATGCTTTTTTCTTTTAAATGACAGCTAGGAACATATGCATCTTTACATGGAAGAA 3138
|||

QY 1443 TCATTGTAAGCTCCTCAATCAACTCTTTAAATCTAAGGSCAACTATGATGAAGGCT 1502
|||
Db 3137 TCATTGTAAGCTCCTCAATCAACTCTTTAAATCTAAGGSCAACTATGATGAAGGCT 3078
|||

QY 1503 TTGGGCACACACACAGGATCTATGGCAAGCAAAAATGAAACGGAAGATTTTGG 1562
|||
```

```
Db 3077 TTGGGCACACACACAGGATCTATGGCAAGCAAAAATGAAACGGAAGATTTTGG 3018
QY 1563 AGAGACCCAGCTTGCAAATGCAAGGAGAGACCCCTCACAGCCAGGGGTAGAGATG 1622
|||
Db 3017 AGAGACCCAGCTTGCAAATGCAAGGAGAGACCCCTCACAGCCAGGGGTAGAGATG 2958
|||
QY 1623 CCCCTATTCTAAGGTGGGTGCTCGGTGCAAGTATGGAAGCCAGGCTCTCTCAGC 1682
|||
Db 2957 CCCCTATTCTAAGGTGGGTGCTCGGTGCAAGTATGGAAGCCAGGCTCTCTCAGC 2898
|||
QY 1683 AGGAGAAGGAAGACCAAGCCAGCTGAAACCAAGAACTGAGGATCGCTCGGCCACCCCA 1742
|||
Db 2897 AGGAGAAGGAAGACCAAGCCAGCTGAAACCAAGAACTGAGGATCGCTCGGCCACCCCA 2838
|||
QY 1743 CTGAATCTGGAAGTTGAGGAAGTGCCTTGGAGGAAGGATCAAAAATGTCAAAAGCCCAAT 1802
|||
Db 2837 CTGAATCTGGAAGTTGAGGAAGTGCCTTGGAGGAAGGATCAAAAATGTCAAAAGCCCAAT 2778
|||
QY 1803 GGCCTCTGGAAGTTGAGGAAGTGCCTTGGAGGAAGGATCAAAAATGTCTAGATCTGA 1862
|||
Db 2777 GGCCTCTGGAAGTTGAGGAAGTGCCTTGGAGGAAGGATCAAAAATGTCTAGATCTGA 2718
|||
QY 1863 AGAAGCTAAGACGATCTTCTCACTGAAGGAAGGAGCGGCCCATTCACGTAGCAGCTT 1922
|||
Db 2717 AGAAGCTAAGACGATCTTCTCACTGAAGGAAGGAGCGGCCCATTCACGTAGCAGCTT 2658
|||
QY 1923 CATTTCAAGCACCTCTGTCAAGAGCCCAAAAATCTGTGTCCCACTATCAAGAAAGGCT 1982
|||
Db 2657 CATTTCAAGCACCTCTGTCAAGAGCCCAAAAATCTGTGTCCCACTATCAAGAAAGGCT 2598
|||
QY 1983 GGAGCATGTGAGAGAGTGAAGAGTGTGTGGTGGAGAGTTCAGAAAAGGAACAAG 2042
|||
Db 2597 GGAGCATGTGAGAGAGTGAAGAGTGTGTGGTGGAGAGTTCAGAAAAGGAACAAG 2538
|||
QY 2043 TGGAAATGCCAAGGCTTCTTAAGAAAGATGGAATGTGGGAAAAACAACCTGGCAAAACA 2102
|||
Db 2537 TGGAAATGCCAAGGCTTCTTAAGAAAGATGGAATGTGGGAAAAACAACCTGGCAAAACA 2478
|||
QY 2103 AAGAATCTAAAGGAGAGACAGGGAAGAGAAGTAAAGAAAGGTCAATAGTTGGAGATGGAGA 2162
|||
Db 2477 AAGAATCTAAAGGAGAGACAGGGAAGAGAAGTAAAGAAAGGTCAATAGTTGGAGATGGAGA 2418
|||
QY 2163 ATGAGAATCTTTAGAAAAATGGTGCAGACTCCGATGAAGATGATAAACAGCTTCCTCAAAAC 2222
|||
Db 2417 ATGAGAATCTTTAGAAAAATGGTGCAGACTCCGATGAAGATGATAAACAGCTTCCTCAAAAC 2358
|||
QY 2223 AACAACTCCACAAGAACCCCAAGTCTCTGAATGGTCCAGTTTTGTAGACAAACACCTTTG 2282
|||
Db 2357 AACAACTCCACAAGAACCCCAAGTCTCTGAATGGTCCAGTTTTGTAGACAAACACCTTTG 2298
|||
QY 2283 CTGAAGAAATTCACCTACTCAGAAATCCAGGATCCAGGATGTGGAATCTGGAGGGAGAG 2342
|||
Db 2297 CTGAAGAAATTCACCTACTCAGAAATCCAGGATCCAGGATGTGGAATCTGGAGGGAGAG 2238
|||
QY 2343 TGGTCAAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATCATGAGGATG 2402
|||
Db 2237 TGGTCAAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGTATTATCATGAGGATG 2178
|||
QY 2403 AGGATGAAGAGTGAACAAATTCGAATGATGTGGGCCCTTTAAATTCATGTTAGTGTAGCGA 2462
|||
Db 2177 AGGATGAAGAGTGAACAAATTCGAATGATGTGGGCCCTTTAAATTCATGTTAGTGTAGCGA 2118
|||
QY 2463 GCCACTGCCCTTTGTCAAAAATGTATGCACATAAGAGAGGTATCCAGCATGAAATGTAAT 2522
|||
Db 2117 GCCACTGCCCTTTGTCAAAAATGTATGCACATAAGAGAGGTATCCAGCATGAAATGTAAT 2058
|||
QY 2523 TTACTTGAAGTAACTTTGGAAGAAAGATTCCTTCTTAAATCAAAAACAAAACAAAAA 2582
|||
Db 2057 TTACTTGAAGTAACTTTGGAAGAAAGATTCCTTCTTAAATCAAAAACAAAACAAAAA 1998
|||
QY 2583 CAAAAAACAACATCTTAAATACTAGAGATAACCTTTTAAATCTTCTTAAATCTTCTTAGCAGT 2642
|||
Db 1997 CAAAAAACAACATCTTAAATACTAGAGATAACCTTTTAAATCTTCTTAAATCTTCTTAGCAGT 1938
|||
```


QY	2643	GATGATATGCATAAGTGTCTGAAGCTTGTAACTGGGGAATATTCCACCTGATTAATAGC	2702
Db	1937	GATGATATCGTAAGTGTCTGAAGCTTGTAACTGGGGAATATTCCACCTGATTAATAGC	1878
QY	2703	CCAGATTCTACTGTATTCCAAAAGGCAATATTAAAGGTAGATAGATGATTAGTAGTATAT	2762
Db	1877	CCAGATTCTACTGTATTCCAAAAGGCAATATTAAAGGTAGATAGATGATTAGTAGTATAT	1818
QY	2763	TGTTACACACTATTTTGGAAATTAGAGAACATACAGAAAGAAATTTAGGGCTTTAAACATTA	2822
Db	1817	TGTTACACACTATTTTGGAAATTAGAGAACATACAGAAAGAAATTTAGGGCTTTAAACATTA	1758
QY	2823	CCACTGAATGCCTTTAGTATTAAGGGCACAGTTTGTATATTTTAAATGAATACCAATT	2882
Db	1757	CCACTGAATGCCTTTAGTATTAAGGGCACAGTTTGTATATTTTAAATGAATACCAATT	1698
QY	2883	TAAATTTTGTAGTATTTACCTGTTTAAGAGATTTATAGTCTTTTAAATTTTGTAGGTTAATT	2942
Db	1697	TAAATTTTGTAGTATTTACCTGTTTAAGAGATTTATAGTCTTTTAAATTTTGTAGGTTAATT	1638
QY	2943	TTCTTGCTGTATATATAGAGAAATTTACTACTTTTATGCTGCTCTCTAAACTACATC	3002
Db	1637	TTCTTGCTGTATATATAGAGAAATTTACTACTTTTATGCTGCTCTCTAAACTACATC	1578
QY	3003	CTGAACCTCAGCTCTGAGGTATATACAAACAGACACTTTTGTAGGGCAATTTGAAAAACC	3062
Db	1577	CTGAACCTCAGCTCTGAGGTATATACAAACAGACACTTTTGTAGGGCAATTTGAAAAACC	1518
QY	3063	AACTACACTCTTCGGTGCTTAGAGAGATCTGCTGCTCCGAAATAAGCTTTTGTATCTG	3122
Db	1517	AACTACACTCTTCGGTGCTTAGAGAGATCTGCTGCTCCGAAATAAGCTTTTGTATCTG	1458
QY	3123	CCAGTGAATTTACTGTACTCCAAATGATTCCTTTCTTGCTGATATCTGTGCTTCTC	3182
Db	1457	CCAGTGAATTTACTGTACTCCAAATGATTCCTTTCTTGCTGATATCTGTGCTTCTC	1398
QY	3183	ATAATTAAGTGAAGTGAATTTTGTAGTAATACCTTCGGGATCACTGTGCCCACTTTC	3242
Db	1397	ATAATTAAGTGAAGTGAATTTTGTAGTAATACCTTCGGGATCACTGTGCCCACTTTC	1338
QY	3243	CTGTGTAGAGCAAGTGAAGGTTTAAAGGAGGAGAGAGAACTGTCTTACACCACT	3302
Db	1337	CTGTGTAGAGCAAGTGAAGGTTTAAAGGAGGAGAGAGAACTGTCTTACACCACT	1278
QY	3303	TGAGCTCAGACCTCTAAACCCCTGATTTCCCTTATGATGTCCTTTTGTAGACACTAAT	3362
Db	1277	TGAGCTCAGACCTCTAAACCCCTGATTTCCCTTATGATGTCCTTTTGTAGACACTAAT	1218
QY	3363	TTTTAAATCTTACTAGCTCTGAAATATATGATTTTATCACAGTATTTCTCAGGGTGAA	3422
Db	1217	TTTTAAATCTTACTAGCTCTGAAATATATGATTTTATCACAGTATTTCTCAGGGTGAA	1158
QY	3423	ATTAACCAACTATAGGCTTTTCTTGGGATGATTTCTAGCTTTAAGGTTTGGGGACA	3482
Db	1157	ATTAACCAACTATAGGCTTTTCTTGGGATGATTTCTAGCTTTAAGGTTTGGGGACA	1098
QY	3483	TTATAAATCTTACTAGTATTTGTACACAGTGTATTTCCAAATTTGATCGATGGGAGG	3542
Db	1097	TTATAAATCTTACTAGTATTTGTACACAGTGTATTTCCAAATTTGATCGATGGGAGG	1038
QY	3543	GAGAGGTGTCTTAAGCTGTAGGCTTTTCTTGTACTGCAATTTATAGATTTAGCTTTAA	3602
Db	1037	GAGAGGTGTCTTAAGCTGTAGGCTTTTCTTGTACTGCAATTTATAGATTTAGCTTTAA	978
QY	3603	TATTTTGTAGAGATGATAAACAATTCGCTTTCTTGTAGTCTTACCTAGTCTGAAACATTTT	3662
Db	977	TATTTTGTAGAGATGATAAACAATTCGCTTTCTTGTAGTCTTACCTAGTCTGAAACATTTT	918
QY	3663	ATTCATAAAGATTTTAAATTAATTTTGA 3692	
Db	917	ATTCATAAAGATTTTAAATTAATTTTGA 888	

RESULT 12	BC001247	2164 bp	mRNA	linear	PR1 12-JUL-2001
LOCUS	BC001247				
DEFINITION	Homo sapiens, epithelial protein lost in neoplasm beta, clone				
	MGC:4969 IMAGE:3452714, mRNA, complete cds.				
ACCESSION	BC001247				
VERSION	BC001247.1	GI:12654808			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2164)				
AUTHORS	Strausberg,R.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-DEC-2000) National Institutes of Health, Mammalian				
	Gene Collection (MGC), Cancer Genomics Office, National Cancer				
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,				
	USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk				
	Email: cgabs-remail.nih.gov				
	Tissue Procurement: ATCC				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Sequencing Group at the Stanford Human Genome				
	Center, Stanford University School of Medicine, Stanford, CA 94305				
	Web site: http://www-shgc.stanford.edu				
	Contact: (Dickson, Mark) mcd@paxil.stanford.edu				
	Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,				
	R. M.				
Clone distribution: MGC clone distribution information can be found					
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov					
Series: IRAX Plate: 4 Row: 0 Column: 9					
This clone was selected for full length sequencing because it					
passed the following selection criteria: matched mRNA gi: 7020350.					
FEATURES	Location/Qualifiers				
	1..2164				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="LocusID:51474"				
	/db_xref="taxon:9606"				
	/clone="MGC:4969 IMAGE:3452714"				
	/tissue_type="Placenta, choriocarcinoma"				
	/clone_lib="NIH MGC 10"				
	/lab_host="DH10B"				
	/note="Vector: pCMV-SPORT6"				
	187..1983				
	/codon_start=1				
	/product="epithelial protein lost in neoplasm beta"				
	/protein_id="AAH01247.1"				
	/db_xref="GI:12654809"				
	/translation="MENCIGSRHVEKSEISENTDASGKIEKYNVPLNRLKMFPEK				
	EPOTKILRAQSRASGRKISNSYSLSLDELEGQOLSSSTFDSKESNRNLELRL				
	SETISKDANKYAAVRSKSSSTNYTNELKASGEIKIHMEQKENVPGPEVCITHQ				
	EGKISDRANSJAVRSTPAEDSRDSQVQVHPKPLSPDSRASSLSSEVCPKA				
	MKFQAPARTCEQKTVYPMERLLANQQVFIHISCFRCYCNKLSLGTYSLSHGRYI				
	CKPHFNQPKSGKNYDEGFRPHKOLWASKNEBELERPAOLANARETPHSPGVED				
	APIAKYGVLAASMEAKASSQEKEDKPAETKRLRIAMPPTTELSSGSALEEGIKMSK				
	PKWPPDETSKEVEPDEVLDLKKRRSSSLKRSRPFTVAQSFQSTVSKPTKVSPP				
	IRKWSMSQSSVEGVRVAERKVENAKSKNGNVKKTWQNKESKGETGKRSKEG				
	HSLMENLVNGADSDDDNFKQQSPQEPKSLNWSFVDNTPAEFTTQNKSQ				
	DVELMEGVVVKELSVBEQIKRNRYYDEDEEE"				
BASE COUNT	744 a	472 c	513 g	435 t	
ORIGIN					
Query Match	57.8%; Score 2141; DB 9; Length 2164;				
Best Local Similarity	99.6%; Pred. No. 0;				
Matches 2157; Conservative	0; Mismatches 5; Indels 3; Gaps 1;				

Db 1 CACAGAGCAGACCATCTCTCTCTGCTGAAAGTACAAGCCACGCTGCTTCTGAGGCCAAGCT 60
QY 491 GACCAAGAGAACAAATCCACCCAGATCTAGACTCAGTCACTCTCTGAAAGCCCTCGTT 550
Db 61 GACCAAGAGAACAAATCCACCCAGATCTAGACTCAGTCACTCTCTGAAAGCCCTCGTT 120
QY 551 CAGGCTCGATATCCCCACATCAAGACCGTGAGGATCTTAAAGACCACTCAACAGABAAGT 610
Db 121 CAGGCTCGATATCCCCACATCAAGACCGTGAGGATCTTAAAGACCACTCAACAGABAAGT 180
QY 611 AAAAATAATGGAAATTTGTCTAGGAGATCCAGGCAATGAAAGTAGAAAAATCAGAAATCAGT 670
Db 181 AAAAATAATGGAAATTTGTCTAGGAGATCCAGGCAATGAAAGTAGAAAAATCAGAAATCAGT 240
QY 671 GAAACACAGATGCTCGGGCAAAATAGAGAAATATATATGTTCCGCTGAAACAGGCTTAAG 730
Db 241 GAAACACAGATGCTCGGGCAAAATAGAGAAATATATATGTTCCGCTGAAACAGGCTTAAG 300
QY 731 ATGATGTTTGAAGAGGTGAACCAACTCAAACTAAGATTTCTCGGGCCCAAAAGCCGAAAT 790
Db 301 ATGATGTTTGAAGAGGTGAACCAACTCAAACTAAGATTTCTCGGGCCCAAAAGCCGAAAT 360
QY 791 GCAAGTGGAAAGAGATCTCTGAAACAGCTATTCTCTAGATGACCTTGGAAATAGGCCCA 850
Db 361 GCAAGTGGAAAGAGATCTCTGAAACAGCTATTCTCTAGATGACCTTGGAAATAGGCCCA 420
QY 851 GGTCAAGTTGATCTTCTAGATTTGATCTCGGAGAAAAATGAGAGTAGACGAAATCTGGAA 910
Db 421 GGTCAAGTTGATCTTCTAGATTTGATCTCGGAGAAAAATGAGAGTAGACGAAATCTGGAA 480
QY 911 CTTCCAGCTCTCAGAAACCTCTATAAGGATCGAATGCCCAAGTACAGGACAGCTGTG 970
Db 481 CTTCCAGCTCTCAGAAACCTCTATAAGGATCGAATGCCCAAGTACAGGACAGCTGTG 540
QY 971 TCCAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGCGGAATC 1030
Db 541 TCCAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGCGGAATC 600
QY 1031 AAAATTCAATAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCCTGAGGTCGATCAAC 1090
Db 601 AAAATTCAATAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCCTGAGGTCGATCAAC 660
QY 1091 CATTGAGAGGGGAAAGATTTCTGCAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCT 1150
Db 661 CATTGAGAGGGGAAAGATTTCTGCAATGAGAAATAGCCTGGCAGTCCGTTCCACCCCT 720
QY 1151 GCCGAAGATGACTCCCGTACTCCAGGTTAAGAGTGAGGTTCAACAGGCTGTCCATCCC 1210
Db 721 GCCGAAGATGACTCCCGTACTCCAGGTTAAGAGTGAGGTTCAACAGGCTGTCCATCCC 780
QY 1211 AAGCCACTAAGTCCAGATCCAGAGGCTCCAGTCTTTCTGAAAGTTCTCTCCCAAGCA 1270
Db 781 AAGCCACTAAGTCCAGATCCAGAGGCTCCAGTCTTTCTGAAAGTTCTCTCCCAAGCA 840
QY 1271 ATGAAGAGTTTCAGGCACTGCAAGAGAGACCTCGGTGGAATGTGAGAGACAGTCTAT 1330
Db 841 AT---GAAAGTTTCAGGCACTGCAAGAGAGACCTCGGTGGAATGTGAGAGACAGTCTAT 897
QY 1331 CCAATGGAGGCTCTGTGGCAACCGAGGAGGTTTTCACATCAGTGTCTTCGTTGCTCC 1390
Db 898 CCAATGGAGGCTCTGTGGCAACCGAGGAGGTTTTCACATCAGTGTCTTCGTTGCTCC 957
QY 1391 TATTGCAACAAACAACTCAGTCTAGGAACATATGCATCTTTACATGGAGAGATCTATTGT 1450
Db 958 TATTGCAACAAACAACTCAGTCTAGGAACATATGCATCTTTACATGGAGAGATCTATTGT 1017
QY 1451 AAGCCTCACTTCAATCAATCTTTTAAATCTAAGGCACTATGATGAAGGCTTTTGGGAC 1510
Db 1018 AAGCCTCACTTCAATCAATCTTTTAAATCTAAGGCACTATGATGAAGGCTTTTGGGAC 1077
QY 1511 AGACCACACAAAGATCTATGGGCAAGCAAAATGAAACGAAAGATTTTGGAGAGACCA 1570

Db 1078 AGACCACACAAAGGATCTATGGGCAAGCAAAATGAAACGAAAGATTTTGGAGAGACCA 1137
QY 1571 GCCAGCTTTGCAAAATGCAAGGAGAGCCCTCACAGCCAGGGGTAGAAGATGCCCTATT 1630
Db 1138 GCCAGCTTTGCAAAATGCAAGGAGAGCCCTCACAGCCAGGGGTAGAAGATGCCCTATT 1197
QY 1631 GCTAAGGTGGGTGCTCTGCTGCTGCAAGTATGGAAGCCAAAGSCCTCTCTCAGCAGGAGAG 1690
Db 1198 GCTAAGGTGGGTGCTCTGCTGCTGCAAGTATGGAAGCCAAAGSCCTCTCTCAGCAGGAGAG 1257
QY 1691 GAAGACAAAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGCTGCCACCCCTCACTGAACTT 1750
Db 1258 GAAGACAAAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGCTGCCACCCCTCACTGAACTT 1317
QY 1751 GGAAGTTTCAGGAAGTGCCTTTGGAGGAAGGATCAAAATGTCAAAGCCCAAAATGGCTCTCT 1810
Db 1318 GGAAGTTTCAGGAAGTGCCTTTGGAGGAAGGATCAAAATGTCAAAGCCCAAAATGGCTCTCT 1377
QY 1811 GAAGACAAATCAGCAAGCCGAAATGCTCTGAGGATGTGATCTAGATCTGAAGAGCTA 1870
Db 1378 GAAGACAAATCAGCAAGCCGAAATGCTCTGAGGATGTGATCTAGATCTGAAGAGCTA 1437
QY 1871 AGACGATCTTCTCACTGAAAGAAAGAGCCGCCCATTCACCTGTAGCAGCTTCATTTCAA 1930
Db 1438 AGACGATCTTCTCACTGAAAGAAAGAGCCGCCCATTCACCTGTAGCAGCTTCATTTCAA 1497
QY 1931 AGACCTCTGTCAAGAGCCCAAAACTGTGTCTCCACACTATCAGGAAAGGCTGGAGCATG 1990
Db 1498 AGACCTCTGTCAAGAGCCCAAAACTGTGTCTCCACACTATCAGGAAAGGCTGGAGCATG 1557
QY 1991 TCAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGAGTTGCAGAAAGGAAACAACTGCAAAAT 2050
Db 1558 TCAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGAGTTGCAGAAAGGAAACAACTGCAAAAT 1617
QY 2051 GCCAAGGCTTCTAAGAGAAATGGGAATGTGGGAAAAAACAACCTGCAAAACAAAGAAATCT 2110
Db 1618 GCCAAGGCTTCTAAGAGAAATGGGAATGTGGGAAAAAACAACCTGCAAAACAAAGAAATCT 1677
QY 2111 AAAGGAGAGACAGGGAAGAGAGTAAGGAAGTCTAGTTGGAGATGGAGATCAGAAAT 2170
Db 1678 AAAGGAGAGACAGGGAAGAGAGTAAGGAAGTCTAGTTGGAGATGGAGATCAGAAAT 1737
QY 2171 CTTGTAGAAAAATGGTGCAGACTCCGATCAAGATGATAACAGCTTCTCTCAAAACAACTCT 2230
Db 1738 CTTGTAGAAAAATGGTGCAGACTCCGATCAAGATGATAACAGCTTCTCTCAAAACAACTCT 1797
QY 2231 CCACAAAGCCCAAGTCTCTGAAATTTGGTGCAGTTTGTAGACAAACACCTTTGCTGAAGAA 2290
Db 1798 CCACAAAGCCCAAGTCTCTGAAATTTGGTGCAGTTTGTAGACAAACACCTTTGCTGAAGAA 1857
QY 2291 TTCCTACTCAGAAATCAGAAATCCAGGATGTGAACTCTCTGGGAGGAGAGAGTGGTCAAA 2350
Db 1858 TTCCTACTCAGAAATCAGAAATCCAGGATGTGAACTCTCTGGGAGGAGAGAGTGGTCAAA 1917
QY 2351 GAGCTCTCTGTGGAGAACACAGATAAGAGAAATCGGTATTATGATGAGGATCAGAGTAA 2410
Db 1918 GAGCTCTCTGTGGAGAACACAGATAAGAGAAATCGGTATTATGATGAGGATCAGAGTAA 1977
QY 2411 GAGTGACAAATTTGCAATCATGCTGGGCTTTAAATTTCAATGTTAGTGTAGCGAGCCTGC 2470
Db 1978 GAGTGACAAATTTGCAATCATGCTGGGCTTTAAATTTCAATGTTAGTGTAGCGAGCCTGC 2037
QY 2471 CCTTTGTCAAAATGTGATGCACATGAGCAGGTATCCAGCATGAAATGTAATTTACTTGG 2530
Db 2038 CCTTTGTCAAAATGTGATGCACATGAGCAGGTATCCAGCATGAAATGTAATTTACTTGG 2097
QY 2531 AAGTAACTTTGGAAAGAAATCTCTTCTTAAATCAAAAAACAAACAAAAACAAAAA 2590
Db 2098 AAGTAACTTTGGAAAGAAATCTCTTCTTAAATCAAAAAACAAACAAAAACAAAAA 2157
QY 2591 ACACA 2595
Db 2158 AAAAA 2162

```

RESULT 13
BC010664          2121 bp      mRNA      linear      PRI 12-JUL-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2121)
Strausberg,R.
Direct Submission
Submitted (10-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalona@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M. Telford, B, Hodgson, A., Bouck, J., Yu, W.,
Murny,D.M., Gibbs,R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC plate: 13 Row: b Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
            Location/Qualifiers
                1..2121
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clones="IMAGE:3854371"
                /tissue_type="Colon, adenocarcinoma"
                /clone_lib="NIH MGC_65"
                /lab_host="DH10B"
                /note="Vector: pCMV-SPORT6"
                <1..831
                /codon_start=1
                /product="Unknown (protein for IMAGE:3854371)"
                /protein_id="AAH10664.1"
                /db_xref="GI:14715009"
                /translations="ARETHPSPGVEDAPIAKVGLAASMEKASSQEKEDPAETKK
                LRIAWPPTELGGSGNLEGIKMSKPKPPPEDEISKVEPVEDVDLDLKKRRSSLK
                ERSRPTVAASFGTSVKSTVSPPIRKGSMSQSEESVGRVAERKQVKNRAKSK
                KNGVOKTITWQNKSEKTKRKGHELMENELVENGASDESDNNSFLKQQSQPE
                PKSLNWSPPVDNTFAEFTTQNKSQDVELWEGEVVKLSVEEQIKRNVYDEDEDE
                "
BASE COUNT      697 a   376 c   448 g   600 t
ORIGIN
Query Match      57.1%; Score 2115.2; DB 9; Length 2121;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2117; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

CY 1586  GCAAGGAGACCCCTCACAGCCAGGGTAGAGATGCCCTATTGCTAAGTGGGTGTC 1645
Db      1  GCAAGGAGACCCCTCACAGCCAGGGTAGAGATGCCCTATTGCTAAGTGGGTGTC 60

```


QY	1248	CTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCG	1307
Db	421	CTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCG	480
QY	1308	TGGAAATGTCAGAGACAGTCTATCCAAATGAGAGGTCTCTTGGCCAAACCACAGAGTGTTC	1367
Db	481	TGGAAATGTCAGAGACAGTCTATRCATGGAGAGGTCTCTTGGCCAAACCACAGAGTGTTC	540
QY	1368	ACATCAGCTGCTTCCGTTGTCTCTATTGCAACCAACCAACTCAGTCTAGGAAACATATGCAT	1427
Db	541	ACATCAGCTGCTTCCGTTGTCTCTATTGCAACCAACCAACTCAGTCTAGGAAACATATGCAT	600
QY	1428	CTTTACATGGAAGAATCTATTGTAAGCCTCACTTTCAATCAACTCTTTAAATCTAAGGGCA	1487
Db	601	CTTTACATGGAAGAATCTATTGTAAGCCTCACTTTCAATCAACTCTTTAAATCTAAGGGCA	660
QY	1488	ACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCTATGGCAAGCAAAAATGAAA	1547
Db	661	ACTATGATGAAGGCTTTGGGCACAGACCACCAAGGATCTATGGCAAGCAAAAATGAAA	720
QY	1548	ACGAAGAGATTTTGGAGAGACCAGCCAGCTTGCAAATGCAAGGGAGACCCCTCAGAGC	1607
Db	721	ACGAAGAGATTTTGGAGAGACCAGCCAGCTTGCAAATGCAAGGGAGACCCCTCAGAGC	780
QY	1608	CAGGGTGAAGATGCCCTATTGCTAAGGTGGGTCTCTGGCTGCAAGTATGGAAGCCA	1667
Db	781	CAGGGTGAAGATGCCCTATTGCTAAGGTGGGTCTCTGGCTGCAAGTATGGAAGCCA	840
QY	1668	AGGCTTCTCTAGCAGGAGAGGAAGACAAGCCAGCTGAAACCAAGAGCTGAGGATCG	1727
Db	841	AGGCTTCTCTAGCAGGAGAGGAAGACAAGCCAGCTGAAACCAAGAGCTGAGGATCG	900
QY	1728	CCTGGCCACCCCTGAGCTTGGAGTTCAGGAAGTTCAGGAAGGATCAAAA	1787
Db	901	CCTGGCCACCCCTGAGCTTGGAGTTCAGGAAGTTCAGGAAGGATCAAAA	960
QY	1788	TGTCAAAGCCCAATGGCTCTCTGAAGAGCAAAATCAGCAAGCCGGAAGTTCCTGAGGATG	1847
Db	961	TGTCAAAGCCCAATGGCTCTCTGAAGAGCAAAATCAGCAAGCCGGAAGTTCCTGAGGATG	1020
QY	1848	TGATCTAGATCTGAAGAAGCTTAAGACGATCTTCTTCACTGAAGGAAGAAGCCGCCAT	1907
Db	1021	TGATCTAGATCTGAAGAAGCTTAAGACGATCTTCTTCACTGAAGGAAGAAGCCGCCAT	1080
QY	1908	TCACCTGTACAGCTTCATTTCAAGCAGCTCTGTCAAGAGCCCAAAAATCTGTGTCCCCAC	1967
Db	1081	TCACCTGTACAGCTTCATTTCAAGCAGCTCTGTCAAGAGCCCAAAAATCTGTGTCCCCAC	1140
QY	1968	CTATCAGGAAGGCTGGAGCATGTCTCAGACAGAGTGAAGAGTCTGTGGTGGCAAGTTG	2027
Db	1141	CTATCAGGAAGGCTGGAGCATGTCTCAGACAGAGTGAAGAGTCTGTGGTGGCAAGTTG	1200
QY	2028	CAGAAAGGAAAAGTGGAAAATGCCAAGGCTTCTTAAGAAGATGGGAATGTGGAAAAA	2087
Db	1201	CAGAAAGGAAAAGTGGAAAATGCCAAGGCTTCTTAAGAAGATGGGAATGTGGAAAAA	1260
QY	2088	CAACCTGGCAAAACAAGAAATCTTAAGGAGACAGAGGAAGAGTGAAGAGGTCTATA	2147
Db	1261	CAACCTGGCAAAACAAGAAATCTTAAGGAGAG-CAGGGAAGAGAGTGAAGAGGTCTATA	1319
QY	2148	GTTTGGAGATGGAGATGAGATCTTTAGAAAAATGGTGCAAGTCTCTGAATTTGGTCGAGTTTG	2207
Db	1320	GTTTGGAGATGGAGATGAGATCTTTAGAAAAATGGTGCAAGTCTCTGAATTTGGTCGAGTTTG	1379
QY	2208	ACAGTCTCTCAAAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCGAGTTTG	2267
Db	1380	ACAGTCTCTCAAAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCGAGTTTG	1439
QY	2268	TAGACAACACCTTTGCTGAAGAAATCTACTACTCAGAAATCCCAAGATGTGGAAC	2327
Db	1440	TAGACAACACCTTTGCTGAAGAAATCTACTACTCAGAAATCCCAAGATGTGGAAC	1499
QY	2328	TCTGGAGGGAGAGTGTCTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGT	2387

Db	1500	TCTGGAGGGAGAGTGGTCAAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGT	1559
QY	2388	ATTATGATGAGGATGAGGATGAGAGTGCACAAATTCGCAATGATGCTGGCCTTAAATTC	2447
Db	1560	ATTATGATGAGGATGAGGATGAGAGTGCACAAATTCGCAATGATGCTGGCCTTAAATTC	1619
QY	2448	TOTTAGTGTAGCGAGCCACTGCCCTTTGTCAAATGTGATGCACATGAAGCAGGTATCCC	2507
Db	1620	TOTTAGTGTAGCGAGCCACTGCCCTTTGTCAAATGTGATGCACATGAAGCAGGTATCCC	1679
QY	2508	AGCATGAATGTAAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTTAAAAATCAA	2567
Db	1680	AGCATGAATGTAAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTTAAAAATCAA	1739
QY	2568	AACAAAAACAAAAA	2582
Db	1740	AAAAAAAACAAAAA	1754
RESULT 15			
BD160117			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
2207 bp DNA linear PAT 17-JAN-2003			
Primer for synthesizing full-length cDNA and use thereof.			
BD160117			
BD160117.1 Gi:27865875			
JP 2002191363-A/14960.			
Homo sapiens (human)			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (Bases 1 to 2207)			
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.			
Primer for synthesizing full-length cDNA and use thereof			
Patent: JP 2002191363-A 14960 09-JUL-2002;			
HELIX RESEARCH INSTITUTE			
OS Homo sapiens (human)			
PN JP 2002191363-A/14960			
PD 09-JUL-2002			
PF 28-JUL-2000 JP 2000280990			
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU			
PI SAITO,			
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,			
PI KEIICHI NAGAI,TETSUJI OTSUKI			
PC			
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/10,			
PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC			
Primer for synthesizing full-length cDNA and use thereof			
FH Key			
Location/Qualifiers			
FT CDS			
Location/Qualifiers			
1..2207			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
BASE COUNT			
ORIGIN			
592 a 438 c 512 g 565 t			
Query Match			
Best Local Similarity 42.3%; Score 1568.4; DB 6; Length 2207;			
Matches 1563; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1001	AATGAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAGGAGAAT	1060
Db	638	AATGAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAGGAGAAT	697
QY	1061	GTCCCCCAGGCTCTGAGGTCTGCATCAACCATCAGGAAGGGGAAAAGATTCTGCAAT	1120
Db	698	GTCCCCCAGGCTCTGAGGTCTGCATCAACCATCAGGAAGGGGAAAAGATTCTGCAAT	757
QY	1121	GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCGAAGATGACTCCCGTACTCCCGGTT	1180

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2004, 09:49:06 ; Search time 903 Seconds
(without alignments)
11075.767 Million cell updates/sec

Title: US-09-890-549-16
Perfect score: 3705
Sequence: 1 ggcgcagcagcagtagtg.....atttgaaaaaaaaaaaaa 3705
Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 252756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq 19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	3705	100.0	3705	21	Sequence encoding
2	3688.4	99.6	3711	21	Human ORFX ORF2698
3	3650.4	98.8	3664	23	Human transcriptio
4	3558.8	96.3	3650	22	Human EPLIN (epith
5	3324.4	89.7	3543	22	Human EPLIN (epith
6	2797.6	75.5	2905	22	Human polynucleoti
7	2789.2	75.3	3465	23	DNA encoding novel
8	2736.4	73.9	2749	24	Human cDNA encodin

9	2696.6	72.8	2783	20	AAZ33566	Human breast tumou
10	2208.8	59.6	2267	21	AAC59489	Human secreted pro
11	2083.6	56.2	2158	21	AAC93483	Human secreted pro
12	1721.4	46.5	1754	24	AB198973	Human cancer suppr
13	1595	43.0	1713	22	AAI59955	Human polynucleoti
14	1568.4	42.3	2207	22	AAH18125	Human cDNA sequenc
15	1433	38.7	1567	21	AAI58588	CDNA encoding huma
16	722	19.5	732	21	AAC58944	Human pancreatic c
17	637.4	17.2	641	20	AAZ52852	Human prostate tum
18	590.6	15.9	698	22	AAH72864	Human cervical can
19	551	14.9	565	24	ABV98930	Human pancreatic c
20	467.4	12.6	1014	23	ABX88277	DNA encoding novel
21	425	11.5	463	24	ABQ58707	Human colon cancer
22	416	11.2	532	22	AAH12979	Human cDNA clone (
23	377.6	10.2	389	21	AAC02079	Human secreted pro
24	362	9.8	586	21	AAA16134	Human colon cancer
25	352.6	9.5	413	25	ACA56082	Chinese hamster si
26	343	9.3	343	24	ABX29495	Colon adenocarcino
27	343	9.3	343	24	ABX29810	Human colon cancer
28	322.2	8.7	382	24	ABQ58853	Human secreted pro
29	321	8.7	376	23	AAH72739	DNA encoding novel
30	294.6	8.0	577	23	ABV49283	Human prostate exp
31	279.8	7.6	283	24	ABV88456	Human colon cancer
32	277.4	7.5	279	25	ACA55895	Chinese hamster si
33	230	6.2	419	22	ABA08387	Human Ca2+ pump PM
34	224	6.0	238	25	ACA55583	Chinese hamster si
35	217.2	5.9	408	23	AAH72740	DNA encoding novel
36	212	5.7	247	22	AAH69368	Human cervical can
37	211.6	5.7	2226	23	AAH72745	DNA encoding novel
38	210.8	5.7	231	22	AAH72234	Human cervical can
39	210.8	5.7	235	22	AAH71170	Human cervical can
40	209.4	5.7	265	22	AAF98667	Human ovarian canc
41	199.2	5.4	296	21	AAC03485	Human secreted pro
42	197.8	5.3	232	22	AAI19758	Human breast cance
43	188.2	5.1	269	22	AAI10178	Human breast tumou
44	175.2	4.7	178	24	ABU38497	Human colon tumour
45	141.2	3.8	2336	22	AAH33341	DNA encoding human

ALIGNMENTS

RESULT 1
AAAS3826
ID AAAS3826 standard; DNA; 3705 BP.
XX
AC AAAS3826;
XX
DT 03-JAN-2001 (first entry)
XX
DE Sequence encoding lipid associated protein (LIPAP) 2766980CBL.
XX
KW Lipid associated protein; LIPAP; treatment; prophylaxis; agonist;
KW antagonist; antibody; cardiovascular disease; neurological disease;
KW gastrointestinal disease; lipid metabolism; detection;
KW amplification; monitoring; hybridisation; antisense; triplex;
KW ribozyme; screening; immunoassay; ds.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 137..2416
FT /*tag= a
FT /product= Lipid associated protein
XX
PN WO200049043-A2.
XX
PD 24-AUG-2000.
XX
PF 18-FEB-2000; 2000WO-US04160.
XX
PR 19-FEB-1999; 99US-0120703.
PR 08-JUL-1999; 99US-0142762.

(INCY-) INCYTE PHARM INC.

Tang YT, Hillman JL, Yue H, Azimzai Y, Baughn MR, Tran B;

WPI: 2000-549264/50.
P-PSDB; AY97286.

New human lipid-associated proteins, nucleic acids, and antibodies, useful for diagnosis, treatment and prevention of e.g. cardiovascular disease

Claim 4; Page 87-88; 93pp; English.

Lipid-associated proteins (LIPAP) can be used for treating or preventing disorders associated with decreased expression of LIPAP, for screening for agonists or antagonists of LIPAP, and to raise specific antibodies. Antagonists and antagonists of LIPAP are useful for treating diseases associated with reduced or increased levels of LIPAP, e.g. cardiovascular, neurological and gastrointestinal diseases and disorders of lipid metabolism. Fragments of the nucleic acid encoding LIPAP are useful for detection of full length coding sequences, in hybridization and/or amplification assays or for diagnosis or monitoring. Nucleotides encoding LIPAP are used to screen for compounds that specifically modify LIPAP expression, for recombinant production of LIPAP, in gene therapy, as a source of therapeutic antisense, triplex-forming, or ribozyme agents and for genomic mapping. Antibodies to the proteins are used for diagnosis and monitoring of LIPAP-associated disease by immunoassay, as antagonists, in competitive drug screens and for affinity purification of natural LIPAP.

Sequence 3705 BP; 1217 A; 743 C; 826 G; 919 T; 0 other;

Query Match	100.0%;	Score 3705;	DB 21;	Length 3705;
Best Local Similarity	100.0%;	Pred No. 0;		
Matches 3705;	Conservative	0; Mismatches	0; Indels	0; Gaps
1	GGCCGAGGACAGTGGTGTAGACGCTTGGTGGCAGAGGTGGGTAGGTAGAGCGCC	60		
1	GGCCGAGGACAGTGGTGTAGACGCTTGGTGGCAGAGGTGGGTAGGTAGAGCGCC	60		
61	GGGACCTGTGACAGGGCTGTAGAGCGCAGAGAAAGGCGGCTTTTAGCCAGGTATTC	120		
61	GGGACCTGTGACAGGGCTGTAGAGCGCAGAGAAAGGCGGCTTTTAGCCAGGTATTC	120		
121	AGTGTCTGTAGCAAGATGGATCATCTCCATTTAATAGACGGCAATGGACCTCACTATC	180		
121	AGTGTCTGTAGCAAGATGGATCATCTCCATTTAATAGACGGCAATGGACCTCACTATC	180		
181	ATTGAGGGTAACAGCCCAAGAACTTTCTTGTCTCAACAGAACTGATCGGCTATTGT	240		
181	ATTGAGGGTAACAGCCCAAGAACTTTCTTGTCTCAACAGAACTGATCGGCTATTGT	240		
241	GGAAATATTCTCCAAGTACCAGAAAGAGCTGAGAGAAACAAATCGGAGAGAGAAG	300		
241	GGAAATATTCTCCAAGTACCAGAAAGAGCTGAGAGAAACAAATCGGAGAGAGAAG	300		
301	TAAACCCGAAATCTCTCCAGCACTTTAGAAAGGGGACCCCTCACTGTGTTAAAGAGAA	360		
301	TAAACCCGAAATCTCTCCAGCACTTTAGAAAGGGGACCCCTCACTGTGTTAAAGAGAA	360		
361	GTGGAGAACCCAGGGCTGGGAGCAGAGTCTCAACAJACTCTCTAGGACAGCAGCAC	420		
361	GTGGAGAACCCAGGGCTGGGAGCAGAGTCTCAACAJACTCTCTAGGACAGCAGCAC	420		
421	TGAGATTAGGCACAGAGCAGACCAATCCCTCTGCTGAAAGTGACACAGCAGCTTCTG	480		
421	TGAGATTAGGCACAGAGCAGACCAATCCCTCTGCTGAAAGTGACACAGCAGCTTCTG	480		
481	AGCCAAAGCTGACCAAGAGAGACAAATCACCCCGAGATCTAGACTCAGGTCACTCTCTGA	540		
481	AGCCAAAGCTGACCAAGAGAGACAAATCACCCCGAGATCTAGACTCAGGTCACTCTCTGA	540		

QY	541	AGCCCTCGTTTCAAGGTCGATATCCCAATCAAGACGGTGAGGATCTTAAAGACCACTC	600
Db	541	AGCCCTCGTTTCAAGGTCGATATCCCAATCAAGACGGTGAGGATCTTAAAGACCACTC	600
QY	601	ACAGAAAGTAAAGAAATGCTAGAGAAATCCAGGCATGAAAGTAGAAAAATC	660
Db	601	ACAGAAAGTAAAGAAATGCTAGAGAAATCCAGGCATGAAAGTAGAAAAATC	660
QY	661	AGAAATCAGTGAAGAAACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAA	720
Db	661	AGAAATCAGTGAAGAAACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAA	720
QY	721	CAGCTTAAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACTAAGATTCTCCGGGCCCA	780
Db	721	CAGCTTAAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACTAAGATTCTCCGGGCCCA	780
QY	781	AAGCCGAAGTGAAGTGAAGAAAGTCTTGAAGACAGCTATTCTCTAGATGACCTGGA	840
Db	781	AAGCCGAAGTGAAGTGAAGAAAGTCTTGAAGACAGCTATTCTCTAGATGACCTGGA	840
QY	841	AATAGGCCAGGTCAAGTGTCTATCTTACATTTTCACTTCGAGAGAAAAATGAGAGTAGACG	900
Db	841	AATAGGCCAGGTCAAGTGTCTATCTTCACTTCGAGAGAAAAATGAGAGTAGACG	900
QY	901	AAATCTGGAACTTCCACGGCTCTCAGAAACCTCTATAAAGGATCGAATGCGCAAGTACCA	960
Db	901	AAATCTGGAACTTCCACGGCTCTCAGAAACCTCTATAAAGGATCGAATGCGCAAGTACCA	960
QY	961	GGCAGCTGTCTCAAAACAAAGCAGCTCAACCACTATACAAATGAGCTGAAGCCAGTGG	1020
Db	961	GGCAGCTGTCTCAAAACAAAGCAGCTCAACCACTATACAAATGAGCTGAAGCCAGTGG	1020
QY	1021	TGGCGAAATCAAAATTCATAAATGGAGGAGAAATGTGCCCCAGGTCTCTGAGGT	1080
Db	1021	TGGCGAAATCAAAATTCATAAATGGAGGAGAAATGTGCCCCAGGTCTCTGAGGT	1080
QY	1081	CTGCATACCCATCAGGAGGGGAAAGATTTCTGCAAAATGAGAAATAGCCTGGAGTCCG	1140
Db	1081	CTGCATACCCATCAGGAGGGGAAAGATTTCTGCAAAATGAGAAATAGCCTGGAGTCCG	1140
QY	1141	TTCCACCCCTGCGGAGAGATGACTCCCGTCACTCCAGGTAAAGAGTGAGTTCAACAGCC	1200
Db	1141	TTCCACCCCTGCGGAGAGATGACTCCCGTCACTCCAGGTAAAGAGTGAGTTCAACAGCC	1200
QY	1201	TGTCCATCCCAAGCCACTAAGTCCAGATCCAGAGCTCCAGTCTTCTGAAAGTTCTCC	1260
Db	1201	TGTCCATCCCAAGCCACTAAGTCCAGATCCAGAGCTCCAGTCTTCTGAAAGTTCTCC	1260
QY	1261	TCGCAAGCAATGAAGAAATTCAGGCACCTGCAAGAGAGACCTCGGTGGAATGTGAGAA	1320
Db	1261	TCGCAAGCAATGAAGAAATTCAGGCACCTGCAAGAGAGACCTCGGTGGAATGTGAGAA	1320
QY	1321	GACAGTCTATCCATGAGGCTCTCTTGGCCAAACCCAGCAGGTGTTTCCACATCAGCTGCTT	1380
Db	1321	GACAGTCTATCCATGAGGCTCTCTTGGCCAAACCCAGCAGGTGTTTCCACATCAGCTGCTT	1380
QY	1381	CGGTGCTCTTATTCGCAACAACTCAATCAATCAATCAATCAATCAATCAATCAATCAAT	1440
Db	1381	CGGTGCTCTTATTCGCAACAACTCAATCAATCAATCAATCAATCAATCAATCAATCAAT	1440
QY	1441	AAATCTATTGTAAGCTCACTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT	1500
Db	1441	AAATCTATTGTAAGCTCACTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT	1500
QY	1501	CTTTGGGACAGACCAACAGGATCTATGGGCAAGCAAAAAATGAAACGAGAGATTTT	1560
Db	1501	CTTTGGGACAGACCAACAGGATCTATGGGCAAGCAAAAAATGAAACGAGAGATTTT	1560
QY	1561	GGAGAGCAGGCGCCAGCTTGCAGAAATCAAGGAGAGACCCCTCACAGCCCGGGGTAGAAGA	1620
Db	1561	GGAGAGCAGGCGCCAGCTTGCAGAAATCAAGGAGAGACCCCTCACAGCCCGGGGTAGAAGA	1620

QY 1621 TGCCCTTATTGCTAAGTGGGTGTCCTGGCTGCAAGTATGGAAGCCAAAGCCCTCCTCTCA 1680
DB 1621 TGCCCTTATTGCTAAGTGGGTGTCCTGGCTGCAAGTATGGAAGCCAAAGCCCTCCTCTCA 1680
QY 1681 GCAGGAGAAGAAACAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGGCCACCCCC 1740
DB 1681 GCAGGAGAAGAAACAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGGCCACCCCC 1740
QY 1741 CACTGAACTTTGGAAGTTTCAGGAAGTGCCTTGGAGGAGGATCAAAATGTCAGAGCCCAA 1800
DB 1741 CACTGAACTTTGGAAGTTTCAGGAAGTGCCTTGGAGGAGGATCAAAATGTCAGAGCCCAA 1800
QY 1801 ATGSCCTTCCTGAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCT 1860
DB 1801 ATGSCCTTCCTGAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCT 1860
QY 1861 GAAGAAGCTAAGACGATCTTCTTCACTGAAGAGAAAGCCGCCCATTCACCTGATAGCAG 1920
DB 1861 GAAGAAGCTAAGACGATCTTCTTCACTGAAGAGAAAGCCGCCCATTCACCTGATAGCAG 1920
QY 1921 TTCAATTTCAAGCACCTCTCTCAAGAGCCCAAAACTGTGTCCCACTATCAGGAAAGG 1980
DB 1921 TTCAATTTCAAGCACCTCTCTCAAGAGCCCAAAACTGTGTCCCACTATCAGGAAAGG 1980
QY 1981 CTGGAGCATCTCAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGAGTTGCCAGAAAGAAACA 2040
DB 1981 CTGGAGCATCTCAGAGCAGAGTGAAGAGTCTGTGGGTGGAAGAGTTGCCAGAAAGAAACA 2040
QY 2041 AGTGGAAATTCGCAAGGCTTCTTAAGAGAAATGGGAATGTGGGAAAAACAACCTGGCAAAA 2100
DB 2041 AGTGGAAATTCGCAAGGCTTCTTAAGAGAAATGGGAATGTGGGAAAAACAACCTGGCAAAA 2100
QY 2101 CAAAGAAATCTAAAGGAGAGACAGGGAAGAGAAATGAAGGAAGTCTAGTGTGGAGATGGA 2160
DB 2101 CAAAGAAATCTAAAGGAGAGACAGGGAAGAGAAATGAAGGAAGTCTAGTGTGGAGATGGA 2160
QY 2161 GAATGAGAAATCTGTAGAAATGCTGAGAAATCCGATGAGATGATAACAGCTTCTCTCAA 2220
DB 2161 GAATGAGAAATCTGTAGAAATGCTGAGAAATCCGATGAGATGATAACAGCTTCTCTCAA 2220
QY 2221 ACAACAATCTCCCAAGAACCCCAAGTCTCTGAATTTGGTTCAGATTTTGTAGACAAACCTT 2280
DB 2221 ACAACAATCTCCCAAGAACCCCAAGTCTCTGAATTTGGTTCAGATTTTGTAGACAAACCTT 2280
QY 2281 TGCTGAAGAAATTCACCTACTCAGAAATCCAGAAATCCAGGATGTGGAATCTCTGGAGGGAGA 2340
DB 2281 TGCTGAAGAAATTCACCTACTCAGAAATCCAGAAATCCAGGATGTGGAATCTCTGGAGGGAGA 2340
QY 2341 AGTGGTCAAGAGCTCTCTGTGGAAGAAACAGATAAAGAGAAATCGGTATTTATGATGAGGA 2400
DB 2341 AGTGGTCAAGAGCTCTCTGTGGAAGAAACAGATAAAGAGAAATCGGTATTTATGATGAGGA 2400
QY 2401 TGAGGATGAAGAGTGAACAATTCGAATGATGTGGGCTTAAATTCATGTTAGTGTAGC 2460
DB 2401 TGAGGATGAAGAGTGAACAATTCGAATGATGTGGGCTTAAATTCATGTTAGTGTAGC 2460
QY 2461 GAGCCACTGCCCCTTTGTCAAAAATGTGATGCACATAAGCAGGTATCCCAAGATGAATGTA 2520
DB 2461 GAGCCACTGCCCCTTTGTCAAAAATGTGATGCACATAAGCAGGTATCCCAAGATGAATGTA 2520
QY 2521 ATTACTTGAAGTAACCTTTGGAAAAAGAAATTCCTTCTTAAATTCAAAAACAACCAAAAA 2580
DB 2521 ATTACTTGAAGTAACCTTTGGAAAAAGAAATTCCTTCTTAAATTCAAAAACAACCAAAAA 2580
QY 2581 AACACAAAAACACATCTTAAATCTAGAGATAACTTTACTTAAATCTTCAATTTAGCA 2640
DB 2581 AACACAAAAACACATCTTAAATCTAGAGATAACTTTACTTAAATCTTCAATTTAGCA 2640
QY 2641 GTGATGATATGCAATAGTGTGTAAGGCTTGAACCTGGGAAATATTCACACCTGATAATA 2700
DB 2641 GTGATGATATGCAATAGTGTGTAAGGCTTGAACCTGGGAAATATTCACACCTGATAATA 2700
QY 2701 GCCAGATTCTACTGTATTCCCAAAAGGCAATATTAAGGTAGATAGATTAGTAGTAT 2760

DB 2701 GCCAGATTCTACTGTATTTCCAAAAGGCAATATTAAAGTAGATAGATTAGTAGTAT 2760
QY 2761 ATTGTTACACACTATTTTGGAAATTAGAGAAACATACAGAAAGAAATTTAGGGCTTTAAACAT 2820
DB 2761 ATTGTTACACACTATTTTGGAAATTAGAGAAACATACAGAAAGAAATTTAGGGCTTTAAACAT 2820
QY 2821 TACGACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATATTTTAAATGAATACCAA 2880
DB 2821 TACGACTGAATGCACCTTTAGTATAAAGGGCACAGTTTGTATATTTTAAATGAATACCAA 2880
QY 2881 TTTAAATTTTTAGTATTTTACCTGTTAAGAGATATTATTAGTCTTTTAAATTTTTAGGTTAA 2940
DB 2881 TTTAAATTTTTAGTATTTTACCTGTTAAGAGATATTATTAGTCTTTTAAATTTTTAGGTTAA 2940
QY 2941 TTTCTTCTGCTGATATATATGAGGAATTTACTACTTTTATGTCCTGCTCTCTAAACTACA 3000
DB 2941 TTTCTTCTGCTGATATATATGAGGAATTTACTACTTTTATGTCCTGCTCTCTAAACTACA 3000
QY 3001 TCCTGAATTCGACGCTCTGAGGTATATAACACAGAGCACCTTTTGGGCAATTTGAAAAA 3060
DB 3001 TCCTGAATTCGACGCTCTGAGGTATATAACACAGAGCACCTTTTGGGCAATTTGAAAAA 3060
QY 3061 CCAACCTACACTCTTCGGTCTTAGAGAGATCTGCTGTCTCCCAATTAAGCTTTTGTATC 3120
DB 3061 CCAACCTACACTCTTCGGTCTTAGAGAGATCTGCTGTCTCCCAATTAAGCTTTTGTATC 3120
QY 3121 TGCCAGTGAATTTACTGTACTCCAAATGATTGCTTCTTTCTGGTGATATCTGTGCTTC 3180
DB 3121 TGCCAGTGAATTTACTGTACTCCAAATGATTGCTTCTTTCTGGTGATATCTGTGCTTC 3180
QY 3181 TCATAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGGATCACTGTCCCCCATCT 3240
DB 3181 TCATAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGGATCACTGTCCCCCATCT 3240
QY 3241 TCCGTGTTAGAGCAAAAGTGAAGAGTTTAAAGAGAGAGAAAGAAAGAACTGTCTTACACA 3300
DB 3241 TCCGTGTTAGAGCAAAAGTGAAGAGTTTAAAGAGAGAGAAAGAAAGAAAGAACTGTCTTACACA 3300
QY 3301 CTGAGCTCAGACCTCTAAACCCCTGATTTCCCTCTATGATGTCCTTTTGTAGACACTA 3360
DB 3301 CTGAGCTCAGACCTCTAAACCCCTGATTTCCCTCTATGATGTCCTTTTGTAGACACTA 3360
QY 3361 ATTTTAAATACTTACTAGCTCTGAAATATATTGATTTTATCAGATTTCTCAGGGTG 3420
DB 3361 ATTTTAAATACTTACTAGCTCTGAAATATATTGATTTTATCAGATTTCTCAGGGTG 3420
QY 3421 AAATTAACCAACTATAGGCTTTTCTTGGGATGATTTTCTAGTCTTAAAGTTTGGGA 3480
DB 3421 AAATTAACCAACTATAGGCTTTTCTTGGGATGATTTTCTAGTCTTAAAGTTTGGGA 3480
QY 3481 CATTATAAATTTGAGTACATTTGTTTACACAGTTGATATTTCCAAATTTGATGATGGGA 3540
DB 3481 CATTATAAATTTGAGTACATTTGTTTACACAGTTGATATTTCCAAATTTGATGATGGGA 3540
QY 3541 GGGAGAGTGTCTTAAGCTGTAGGCTTTTCTTCTAGTCTTACGATTTATAGAGATTTAGCTTT 3600
DB 3541 GGGAGAGTGTCTTAAGCTGTAGGCTTTTCTTCTAGTCTTACGATTTATAGAGATTTAGCTTT 3600
QY 3601 AATATTTTTAGAGATGTAAACCATTTCTGCTTCTTCTAGTCTTACCTAGTCTGAAACATTT 3660
DB 3601 AATATTTTTAGAGATGTAAACCATTTCTGCTTCTTCTAGTCTTACCTAGTCTGAAACATTT 3660
QY 3661 TTATTCATAAAGATTTTAAATTAATAATTTGAAAAAATTTGAAAAAATTTGAAAAAATTT 3705
DB 3661 TTATTCATAAAGATTTTAAATTAATAATTTGAAAAAATTTGAAAAAATTTGAAAAAATTT 3705

RESULT 2
AAC77143
ID AAC77143 standard; cDNA; 3711 bp.
XX
AC AAC77143;

1	GGCCGACAGAGCAGTAGTGGTTAGCAGCTTGGTCCGCACAGGTGGCGCTAGGTAGAGCGCC	60
6	GGCCGACAGAGCAGTAGTGGTTAGCAGCTTGGTCCGCACAGGTGGCGCTAGGTAGAGCGCC	65
61	GGGACCTGTACAGGCGCTGGTAGCAGCGACAGAGAAAGCGCGCTTTTAGCCAGGTATTTTC	120
66	GGGACCTGTACAGGCGCTGGTAGCAGCGACAGAGAAAGCGCGCTTTTAGCCAGGTATTTTC	125
121	AGTGTCTGTAGACAAGATGGAATCATCTCCATTTAATAGACGCAATGGACCTTCACATATC	180
126	AGTGTCTGTAGACAAGATGGAATCATCTCCATTTAATAGACGCAATGGACCTTCACATATC	185
181	ATTGAGGTAACAGCCAAAGAACTTTCTTTGTCAACAAGAACTGATCATCTGCTATGTT	240
186	ATTGAGGTAACAGCCAAAGAACTTTCTTTGTCAACAAGAACTGATCATCTGCTATGTT	245
241	GGAAATATTTCTCAAGTACAGAGAGCAGCTGAAGAAACAAACATGGAGAAGAGAGAG	300
246	GGAAATATTTCTCAAGTACAGAGAGCAGCTGAAGAAACAAACATGGAGAAGAGAGAG	305
301	TAAACCCGAAATCTCTCCAGCACTTTAGAAAGGGGACCTGACTGTGTTAAACAAGAA	360
306	TAAACCCGAAATCTCTCCAGCACTTTAGAAAGGGGACCTGACTGTGTTAAACAAGAA	365
361	GTGGGAGAACCCAGGCGTGGGAGCAGTCTCACACAGACTCTACGGAACAGCAGCAC	420
366	GTGGGAGAACCCAGGCGTGGGAGCAGTCTCACACAGACTCTACGGAACAGCAGCAC	425
421	TGAGATTAGGCAACAGAGCAGACCATCTCTCTGCTGAAGTGACAAGGCCACGCTCTCTGG	480
426	TGAGATTAGGCAACAGAGCAGACCATCTCTCTGCTGAAGTGACAAGGCCACGCTCTCTGG	485
481	AGCCAAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGACTCAGGTCACCTCCTGA	540
486	AGCCAAAGCTGACCAAGAGAAACAAATCCACCCAGATCTAGACTCAGGTCACCTCCTGA	545
541	AGCCCTCGTTCAGGTCGATATCCCCACATCAAGGACGGTGAAGTGAAGAACCACTC	600
546	AGCCCTCGTTCAGGTCGATATCCCCACATCAAGGACGGTGAAGTGAAGAACCACTC	605
601	AACAGAAAGTAAAAAATGGAATTTGTCTAGGAGAAATCCAGGCAATGAAGTAGAAAAATC	660
606	AACAGAAAGTAAAAAATGGAATTTGTCTAGGAGAAATCCAGGCAATGAAGTAGAAAAATC	665
661	AGAAATCAGTGAACACAGATGCTTTCCGGCAAAATAGAGAAATATAATGTTCCGCTGAA	720
666	AGAAATCAGTGAACACAGATGCTTTCCGGCAAAATAGAGAAATATAATGTTCCGCTGAA	725
721	CAGGCTTAAGATGATGTTTGAGAAAGGTGAACCACTCAAACTTAAGATTTCTCCGGGCCA	780
726	CAGGCTTAAGATGATGTTTGAGAAAGGTGAACCACTCAAACTTAAGATTTCTCCGGGCCA	785
781	AAGCCGAAGTGAAGTGAAGGAGAGTCTCTGAAACAGCTATTCTCTAGATGACCTGGA	840
786	AAGCCGAAGTGAAGTGAAGGAGAGTCTCTGAAACAGCTATTCTCTAGATGACCTGGA	845
841	AATAGGCCCAGGTCAAGTTGTCATCTTACATTTGACTCCGAGAAAAATGAGAGTAGACG	900
846	AATAGGCCCAGGTCAAGTTGTCATCTTACATTTGACTCCGAGAAAAATGAGAGTAGACG	905
901	AAATCTGAACTTCCAGCCTCTCAGAAACCTCTATAAAGGATGGAATGGCCAGTACCA	960
906	AAATCTGAACTTCCAGCCTCTCAGAAACCTCTATAAAGGATGGAATGGCCAGTACCA	965
961	GGCAGTGTGTCCAAACAAAGCAGCTCAACCAATATACAAATGAGCTGAAAGCCAGTGG	1020
966	GGCAGTGTGTCCAAACAAAGCAGCTCAACCAATATATACAAATGAGCTGAAAGCCAGTGG	1025
1021	TGGCGAAATCAAAATTCATAAATGGAGCAAGGAGAAATGTGCCCCAGGTCCTTGAGT	1080
1026	TGGCGAAATCAAAATTCATAAATGGAGCAAGGAGAAATGTGCCCCAGGTCCTTGAGT	1085

QY 1081 CTGCATCACCCATCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCCTGGCAGTCGG 1140
Db 1086 CTGCATCACCCATCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCCTGGCAGTCGG 1145
QY 1141 TTCCACCCCTGCCGAAGATCACTCCC...GTGACTCCCAGGTTAAGAGTGAGGTTCAACA 1197
Db 1146 TTCCACCCCTGCCGAAGATCACTCCCAGGTCGACTCCCAGGTTAAGAGTGAGGTTCAACA 1205
QY 1198 GCCTGTCCATCCCAAGCCACTAGTCCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTTC 1257
Db 1206 GCCTGTCCATCCCAAGCCACTAGTCCAGATTCAGAGCCTCCAGTCTTTCTGAAAGTTC 1265
QY 1258 TCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGGTGGAAATGTCA 1317
Db 1266 TCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGGTGGAAATGTCA 1325
QY 1318 GAAGACAGTCTATCCCAATGAGAGCTCTCTTGGCCAAACCAGCAGGTGTTTCACATCAGCTG 1377
Db 1326 GAAGACAGTCTATCCCAATGAGAGCTCTCTTGGCCAAACCAGCAGGTGTTTCACATCAGCTG 1385
QY 1378 CTTCCGTTGCTCTATTGCAACCAACAACTCAGTCTAGGAACATATGCACTCTTTACATGG 1437
Db 1386 CTTCCGTTGCTCTATTGCAACCAACAACTCAGTCTAGGAACATATGCACTCTTTACATGG 1445
QY 1438 AAGAACTATTGTAAGCCTCACCTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGA 1497
Db 1446 AAGAACTATTGTAAGCCTCACCTTCAATCAACTCTTTAAATCTAAGGGCAACTATGATGA 1505
QY 1498 AGGCTTTGGGCACAGCACACAAGAGTCTATGGGCAAGCAAAATGAAAAACGAAGAGAT 1557
Db 1506 AGGCTTTGGGCACAGCACACAAGAGTCTATGGGCAAGCAAAATGAAAAACGAAGAGAT 1565
QY 1558 TTTGGAGAGACCGCCAGCTTGCAAAATGCAAGGAGAGACCCCTCACAGCCCAAGGGGTAGA 1617
Db 1566 TTTGGAGAGACCGCCAGCTTGCAAAATGCAAGGAGAGACCCCTCACAGCCCAAGGGGTAGA 1625
QY 1618 AGATGCCCTATTGCTAAGTGGGTGCTGGTGGTGAAGTATGGAAGCCAAAGCCTCTC 1677
Db 1626 AGATGCCCTATTGCTAAGTGGGTGCTGGTGGTGAAGTATGGAAGCCAAAGCCTCTC 1685
QY 1678 TCAGCAGGAGAGAGAGACAGCCAGCTGAAACCAAGAGCTGAGGATCGCCTGGCCACC 1737
Db 1686 TCAGCAGGAGAGAGAGACAGCCAGCTGAAACCAAGAGCTGAGGATCGCCTGGCCACC 1745
QY 1738 CCCACTGAACTTTGGAAGTTTCAGGAAGTGCCTTTGAGGAAGGATCAAAATGTCAAAGCC 1797
Db 1746 CCCACTGAACTTTGGAAGTTTCAGGAAGTGCCTTTGAGGAAGGATCAAAATGTCAAAGCC 1805
QY 1798 CAAATGGCCTCTCGAAGACCGAAATCAGCAAGCCGAAAGTTCCTGAGGATGTCGATCTAGA 1857
Db 1806 CAAATGGCCTCTCGAAGACCGAAATCAGCAAGCCGAAAGTTCCTGAGGATGTCGATCTAGA 1865
QY 1858 TCTGAAGAGCTAAGAGCATCTTTCTCACTGAAGGAAGAAAGCCGCCAATCACTGTAGC 1917
Db 1866 TCTGAAGAGCTAAGAGCATCTTTCTCACTGAAGGAAGAAAGCCGCCAATCACTGTAGC 1925
QY 1918 AGCTTCATTTCAAAGCACCTCTGTCAGAGCCCAAAAATGTGTCCCCACCTATCAGGAA 1977
Db 1926 AGCTTCATTTCAAAGCACCTCTGTCAGAGCCCAAAAATGTGTCCCCACCTATCAGGAA 1985
QY 1978 AGCCTGGAGCATGTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAAGAGTTGCAGAAAGGAA 2037
Db 1986 AGCCTGGAGCATGTCAGAGCAGAGTGAAGAGTCTGTGGTGGAAAGAGTTGCAGAAAGGAA 2045
QY 2038 ACNAGTGGAAATGCCAAGCTTCTAAGAAAGAAATGGGAATGTGGGAAAAACACCTGGCA 2097
Db 2046 ACNAGTGGAAATGCCAAGCTTCTAAGAAAGAAATGGGAATGTGGGAAAAACACCTGGCA 2105
QY 2098 AAACAAGAAATCTAAGGAGAGACAGGAGAGAGTGAAGAGTCAAGTCTAGTTTGGAGAT 2157
Db 2106 AAACAAGAAATCTAAGGAGAGACAGGAGAGAGTGAAGAGTCAAGTCTAGTTTGGAGAT 2165
QY 2158 GGAGAATGAGAATCTTTGTAGNAATTTGGTGCAGACTCCGATGAAGATGATPACAGCTTCT 2217

Db 2166 GGAGAATGAGAATCTTTGTAGNAATTTGGTGCAGACTCCGATGAAGATGATAACAGCTTCT 2225
QY 2218 CAAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCGAGTTTGTAGACAACAC 2277
Db 2226 CAAACAACAATCTCCACAAGAACCCCAAGTCTCTGAATTTGGTCGAGTTTGTAGACAACAC 2285
QY 2278 CTTTCTCTGAAGAAATTCACCTACTCAGAAATCAGAAATCCCAAGATGTGGAACTCTGGGAGG 2337
Db 2286 CTTTCTCTGAAGAAATTCACCTACTCAGAAATCAGAAATCCCAAGATGTGGAACTCTGGGAGG 2345
QY 2338 AGAAGTGTCTCAAGAGCTCTCTGTGGAAAGACAGATAAAGAGAAATCGGTATTATGATGA 2397
Db 2346 AGAAGTGTCTCAAGAGCTCTCTGTGGAAAGACAGATAAAGAGAAATCGGTATTATGATGA 2405
QY 2398 GGATCAGGATGAAGAGTGCACAAATTTGCAATGATGCTGGCCCTTAAATTCATGTTAGTGT 2457
Db 2406 GGATCAGGATGAAGAGTGCACAAATTTGCAATGATGCTGGCCCTTAAATTCATGTTAGTGT 2465
QY 2458 AGCGAGCCACTGCCCTTTGTCAAAAATGTCATTAAGCAGGTATCCCAAGCATGAAAT 2517
Db 2466 AGCGAGCCACTGCCCTTTGTCAAAAATGTCATTAAGCAGGTATCCCAAGCATGAAAT 2525
QY 2518 GTAAATTTACTTTGGAAAGTAACTTTGGAAAAAGAAATTCCTTTTAAAAATCAAAAAACAA 2577
Db 2526 GTAAATTTACTTTGGAAAGTAACTTTGGAAAAAGAAATTCCTTTTAAAAATCAAAAAACAA 2585
QY 2578 AAAAAACAAAAAACAACATCTTAATCTAGAGATACTTTACTTAAATTTCTTCATTTTA 2637
Db 2586 AAAAAACAAAAAACAACATCTTAATCTAGAGATACTTTACTTAAATTTCTTCATTTTA 2645
QY 2638 GCAGTGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2697
Db 2646 GCAGTGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2705
QY 2698 ATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAGGTAGATAGATGATGATGAT 2757
Db 2706 ATAGCCAGATTTCTACTGTATTTCCCAAAAGGCAATATTAAGGTAGATAGATGATGATGAT 2765
QY 2758 TATATTGTTACACACTATTTTGGAAATTAGAGAAACATACAGAAAGAAATTTAGGGCTTAAA 2817
Db 2766 TATATTGTTACACACTATTTTGGAAATTAGAGAAACATACAGAAAGAAATTTAGGGCTTAAA 2825
QY 2818 CATTACGACTCAATGTCACCTTTAGTATTAAGGGCAGTTTCTATATTTTAAATGAATAC 2877
Db 2826 CATTACGACTCAATGTCACCTTTAGTATTAAGGGCAGTTTCTATATTTTAAATGAATAC 2885
QY 2878 CAATTTAAATTTTGTAGTATTTACCTGTTAAGAGATTTTACTCTTTTAAATTTTGTAGT 2937
Db 2886 CAATTTAAATTTTGTAGTATTTACCTGTTAAGAGATTTTACTCTTTTAAATTTTGTAGT 2945
QY 2938 TAAATTTCTGCTGTATATATGAGAAATTTACTACTTTTATGTCCTGCTCTCTAAACT 2997
Db 2946 TAAATTTCTGCTGTATATATGAGAAATTTACTACTTTTATGTCCTGCTCTCTAAACT 3005
QY 2998 ACATCTCGAACTCCAGCTCTCGAGTATATACAAAGAGACACTTTTGTAGGCAATGAA 3057
Db 3006 ACATCTCGAACTCCAGCTCTCGAGTATATACAAAGAGACACTTTTGTAGGCAATGAA 3065
QY 3058 AAAACCAACTACACTCTTTGGGTGCTTAGAGAGATCTCTGCTCTCCCAAAATAGCTTTGT 3117
Db 3066 AAAACCAACTACACTCTTTGGGTGCTTAGAGAGATCTCTGCTCTCCCAAAATAGCTTTGT 3125
QY 3118 ATCTGCCAGTGAATTTACTGTACTCCAAATGATTTCTTTTCTGGTGAATCTGTGC 3177
Db 3126 ATCTGCCAGTGAATTTACTGTACTCCAAATGATTTCTTTTCTGGTGAATCTGTGC 3185
QY 3178 TTCTCATTAATTTACTGAAGTGCATAATTTTGTAGTAATACCTTCGGGATCACTGTCCCCA 3237
Db 3186 TTCTCATTAATTTACTGAAGTGCATAATTTTGTAGTAATACCTTCGGGATCACTGTCCCCA 3245
QY 3238 TCTTCCGTGTTAGAGCAAAAGTGAAGATTTTAAAGGAGGAAGAAAGAACTGTCTTTACA 3297

3246 TCTTCCTGTTAGACAAAGTGAAGAGTTTAAAGGAGGAGAAAGAAAGAACTGCTTTACA 3305
3298 CCACITGAGCTCAGACCTCTAAGACCTGTATTCCTCTTATGATGTCCCTTTTGTAGACA 3357
3306 CCACITGAGCTCAGACCTCTAAGACCTGTATTCCTCTTATGATGTCCCTTTTGTAGACA 3365
3358 CTAATTTTAAATACCTAGCTCTGAAATATATTGATTTTATCAGATTTCTCAGG 3417
3366 CTAATTTTAAATACCTAGCTCTGAAATATATTGATTTTATCAGATTTCTCAGG 3425
3418 GTGAAATTAACCAACTATAGGCTTTTCTGGGATGATTTCTAGTCTTAAAGTTTGG 3477
3426 GTGAAATTAACCAACTATAGGCTTTTCTGGGATGATTTCTAGTCTTAAAGTTTGG 3485
3478 GGACATTAATAACTTGAGTACATTTGTTAGACAGTTGATTTCCAAATTTGATGGATG 3537
3486 GGACATTAATAACTTGAGTACATTTGTTAGACAGTTGATTTCCAAATTTGATGGATG 3545
3538 GGAGGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTTGACTGCTATTTATAGAGATTAGC 3597
3546 GGAGGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTTGACTGCTATTTATAGAGATTAGC 3605
3598 TTTATATATTTTACAGATGTAACACATTTCTGCTTTCTAGTCTTACCTAGTCTGAAACA 3657
3606 TTTATATATTTTACAGATGTAACACATTTCTGCTTTCTAGTCTTACCTAGTCTGAAACA 3665
3658 TTTTATTTCAATAAGATTTTAAATAAATTTGAAAAAATAAAAA 3703
3666 TTTTATTTCAATAAGATTTTAAATAAATTTGAAAAAATAAAAA 3711

SULT 3
X71307

ABX71307 standard; cDNA; 3664 BP.

ABX71307;

14-APR-2003 (first entry)

Human transcription factor cDNA from clone DKFZphuel_18c19.

Human; gene; gene therapy; vaccine; disease treatment; detection; ss.

Homo sapiens.

WO200112659-A2.

22-FEB-2001.

18-AUG-2000; 2000WO-IB01496.

18-AUG-1999; 99US-0149499.

28-SEP-1999; 99US-0156503.

(GEHU-) GERMAN HUMAN GENOME PROJECT.

Wiemann S;

WPI; 2001-327840/34.

P-PSDB; ABUS2859.

Nucleic acids having the sequences of clones isolated from libraries of

different human tissues, useful in recombinant DNA methodologies -

Claim 1; Page 449-450; 1095pp; English.

This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of

the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence encodes a polypeptide described in the disclosure of the invention.

Sequence 3664 BP; 1210 A; 734 C; 809 G; 911 T; 0 other;

Query March 98.8%; Score 3660.4; DB 23; Length 3664;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3661; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

44 CGCTAGTAGAGCGCGGACCTGTGACAGGGCTGTGACAGCGCAGAGAAAGCGGC 103

1 GCGCTAGTAGAGCGCGGACCTGTGACAGGGCTGTGACAGCGCAGAGAAAGCGGC 60

104 TTTTAGCAGGATTTTCACTGTGTGACAGAGATGGAATCATCTCCATTTAATAGACGG 163

61 TTTTAGCAGGATTTTCACTGTGTGACAGAGATGGAATCATCTCCATTTAATAGACGG 120

164 CAATGGACCTCACTATCATTGAGGGTAACAGCAAGAACTTTCTTGTCAACAAGAAC 223

121 CAATGGACCTCACTATCATTGAGGGTAACAGCAAGAACTTTCTTGTCAACAAGAAC 180

224 AAGTCATCGGCTATTGTGAAATATTCTCAAGTACCAGAAAGCAGCTGAAGAAACAAAC 283

181 AAGTCATCGGCTATTGTGAAATATTCTCAAGTACCAGAAAGCAGCTGAAGAAACAAAC 240

284 ATGGAGAAGAGAGATTAACACCGAAATCTCTCCAGCACCTTTAGAAAGGGGACCTTG 343

241 ATGGAGAAGAGAGATTAACACCGAAATCTCTCCAGCACCTTTAGAAAGGGGACCTTG 300

344 ACTGTGTTAAGAGAGAGTGGGAGAACCCAGGGCTGGGAGAGCTCTCACAGAGCTCT 403

301 ACTGTGTTAAGAGAGAGTGGGAGAACCCAGGGCTGGGAGAGCTCTCACAGAGCTCT 360

404 CTACGGAACAGCAGCACTGAGATTAGGCACAGAGACCACTCTCTCTGCTGAAGTGACA 463

361 CTACGGAACAGCAGCACTGAGATTAGGCACAGAGACCACTCTCTCTGCTGAAGTGACA 420

464 AGCCACGCTGCTCTGGAGCCAAAGCTGACCAAGAAAGAAACAAATCCACCCAGATCTAGA 523

421 AGCCACGCTGCTCTGGAGCCAAAGCTGACCAAGAAAGAAACAAATCCACCCAGATCTAGA 480

524 CTCAGTCACTCTGGAAGCCCTGTTTCAAGGTCGATATCCCAATCAAGACCGGTGAG 583

481 CTCAGTCACTCTGGAAGCCCTGTTTCAAGGTCGATATCCCAATCAAGACCGGTGAG 540

584 GATCTTAAAGACCACTCAACAGAAAGTAAAAAATGGAAAAATGTTCTAGGAGAAATCCAGG 643

541 GATCTTAAAGACCACTCAACAGAAAGTAAAAAATGGAAAAATGTTCTAGGAGAAATCCAGG 600

644 CATGAAGTAGAAAAATCAGAAATCAGTGAACACAGATGCTTCGGGCAAAATAGAGAAA 703

601 CATGAAGTAGAAAAATCAGAAATCAGTGAACACAGATGCTTCGGGCAAAATAGAGAAA 660

704 TATAATGTTCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACT 763

661 TATAATGTTCCGCTGAACAGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACT 720

764 AAGATTCTCCGGGCCCAAGCCGAAGTGCMAAGTGAAGGAAAGATCTCTGAAAAACAGCTAT 823

721 AAGATTCTCCGGGCCCAAGCCGAAGTGCMAAGTGAAGGAAAGATCTCTGAAAAACAGCTAT 780

824 TCTCTAGATGACCTGGAATAGGCCCGAGGTGATGTTGATCTTCTACATTTGACTCGGAG 883

781 TCTCTAGATGACCTGGAATAGGCCCGAGGTGATGTTGATCTTCTACATTTGACTCGGAG 840

884 AAAAATGAGAGTAGCGAAATCTGAACTTCCAGCCCTCTCAGAAACCTCTATAAAGGAT 943

841 AAAAATGAGAGTAGCGAAATCTGAACTTCCAGCCCTCTCAGAAACCTCTATAAAGGAT 900

944 CGAATGGCAAGTACAGGAGCTGTGTCACAAACAAAGCAGCTCAACCAACTATACAAT 1003

901 CGAATGGCAAGTACAGGAGCTGTGTCACAAACAAAGCAGCTCAACCAACTATACAAT 960

QY 1004 GAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTG 1063
DB 961 GAGCTGAAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTG 1020
QY 1064 CCCCCAGGTCTCTGAGTCTGCATCACCCATCAGCAAGGGGAAAGATTTCTGCAAAATGAG 1123
DB 1021 CCCCCAGGTCTCTGAGTCTGCATCACCCATCAGCAAGGGGAAAGATTTCTGCAAAATGAG 1080
QY 1124 AATAGCTGCGAGTCCGTCTCCAGCCCTGCCGAAGATGACTCCCGTGACTCCCGAGTTAAG 1183
DB 1081 AATAGCTGCGAGTCCGTCTCCAGCCCTGCCGAAGATGACTCCCGTGACTCCCGAGTTAAG 1140
QY 1184 AGTGAGGTTCAACAGCGCTGTCATCCCAAGCCACTAAGTCCAGATCCAGAGGCTCCAGT 1243
DB 1141 AGTGAGGTTCAACAGCGCTGTCATCCCAAGCCACTAAGTCCAGATCCAGAGGCTCCAGT 1200
QY 1244 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAATTTTCAGGCACCTGCAAGAGAGACC 1303
DB 1201 CTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAATTTTCAGGCACCTGCAAGAGAGACC 1260
QY 1304 TGGGTGGAATGTGAGAAGACAGTCTATCCAAATGGAGCGTCTCTTGCCCAACCCAGCAGGTG 1363
DB 1261 TGGGTGGAATGTGAGAAGACAGTCTATCCAAATGGAGCGTCTCTTGCCCAACCCAGCAGGTG 1320
QY 1364 TTTCACATCAGCTGCTTCCGTGCTCTCTATTGCAACAACAACTCAGTCTAGGAACATAT 1423
DB 1321 TTTCACATCAGCTGCTTCCGTGCTCTCTATTGCAACAACAACTCAGTCTAGGAACATAT 1380
QY 1424 GCATCTTTACATGGAAGAAATCTATTGTAGCCCTCACTTCAATCAACTCTTTAAATCTAAG 1483
DB 1381 GCATCTTTACATGGAAGAAATCTATTGTAGCCCTCACTTCAATCAACTCTTTAAATCTAAG 1440
QY 1484 GGCAACTATGATGAAGGCTTTGGGCACAGACCCACACAAGGATCTATGGGCACGCAAAAAAT 1543
DB 1441 GGCAACTATGATGAAGGCTTTGGGCACAGACCCACACAAGGATCTATGGGCACGCAAAAAAT 1500
QY 1544 GAAAAAGAGAGATTTTGGAGACACAGCCAGCTTGCAAAATGCAAGGAGACCCCTCAC 1603
DB 1501 GAAAAAGAGAGATTTTGGAGACACAGCCAGCTTGCAAAATGCAAGGAGACCCCTCAC 1560
QY 1604 AGCCAGGGGTAGAAGATGCCCTATTGTGAAGTGGGTGCTCGTGCCTCAAGTATGGA 1663
DB 1561 AGCCAGGGGTAGAAGATGCCCTATTGTGAAGTGGGTGCTCGTGCCTCAAGTATGGA 1620
QY 1664 GCCAAGGCTCTCTCAGCAGGAGAGAGGAGACAAGCCAGCTGAAACCAGAGGCTGAGG 1723
DB 1621 GCCAAGGCTCTCTCAGCAGGAGAGAGGAGACAAGCCAGCTGAAACCAGAGGCTGAGG 1680
QY 1724 ATGCTGCGCCACCCCCCACTGAACTTGAAGTTCAGGAAGTGCCTTGAGGAAGGATC 1783
DB 1681 ATGCTGCGCCACCCCCCACTGAACTTGAAGTTCAGGAAGTGCCTTGAGGAAGGATC 1740
QY 1784 ABAATGTCAGAGCCCAATGGCCCTCCTGAAGAGCAAAATCAGCAGCCCGAAGTCTCTGAG 1843
DB 1741 ABAATGTCAGAGCCCAATGGCCCTCCTGAAGAGCAAAATCAGCAGCCCGAAGTCTCTGAG 1800
QY 1844 GATGTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTCACTGAAGGAGAAAGAGCCGC 1903
DB 1801 GATGTCGATCTAGATCTGAAGAGCTTAAGACGATCTTCTCACTGAAGGAGAAAGAGCCGC 1860
QY 1904 CCATTCACTGTAGCAGCTTCATTTCAAGAGCACTCTGCAAGAGCCCAAAAACTGTGCC 1963
DB 1861 CCATTCACTGTAGCAGCTTCATTTCAAGAGCACTCTGCAAGAGCCCAAAAACTGTGCC 1920
QY 1964 CCACCTATCAGNAAGGCTGGAGCATGTGAGAGCAGAGTGAAGATCTGTGGTGGGAAGA 2023
DB 1921 CCACCTATCAGNAAGGCTGGAGCATGTGAGAGCAGAGTGAAGATCTGTGGTGGGAAGA 1980
QY 2024 GTTGCAAGAGGAAACAAAGTGGAAATGCCAGGCTTCTAAGAGNAATGGGAATGTGGA 2083
DB 1981 GTTGCAAGAGGAAACAAAGTGGAAATGCCAGGCTTCTAAGAGNAATGGGAATGTGGA 2040

QY 2084 AAAACAACTGGCAAAAACAAAGAAATCTTAAAGAGAGACAGGAAAGAGAAAGT 2143
DB 2041 AAAACAACTGGCAAAAACAAAGAAATCTTAAAGAGAGACAGGAAAGAGAAAGT 2100
QY 2144 CATAGTTTGGAGATGGAGAAATCAGAAATCTTGTAGAAAAATGGTGACACTCCGATGAAGAT 2203
DB 2101 CATAGTTTGGAGATGGAGAAATCAGAAATCTTGTAGAAAAATGGTGACACTCCGATGAAGAT 2160
QY 2204 GATAACAGCTTCTCTCAACAAACAAATCTCCACAAAGCCCAAGTCTCTGAAATTTGGTTCGAGT 2263
DB 2161 GATAACAGCTTCTCTCAACAAACAAATCTCCACAAAGCCCAAGTCTCTGAAATTTGGTTCGAGT 2220
QY 2264 TTTGTAGACAACACCTTTGCTGAGAATTTCACTACTCAGAATCAGAAATCCAGGATGTG 2323
DB 2221 TTTGTAGACAACACCTTTGCTGAGAATTTCACTACTCAGAATCAGAAATCCAGGATGTG 2280
QY 2324 GAACTCTCGGAGGAGAGAGTGGTCAACAGAGTCTCTGTGGAAAGAACACAGATAAAGAGAAAT 2383
DB 2281 GAACTCTCGGAGGAGAGAGTGGTCAACAGAGTCTCTGTGGAAAGAACACAGATAAAGAGAAAT 2340
QY 2384 CGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTTGCAATGATGCTGGGCTTTAAA 2443
DB 2341 CGGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTTGCAATGATGCTGGGCTTTAAA 2400
QY 2444 TTTCATGTTAGTGTAGCCAGCCACTGCCCTTTGTCCAAATGTGTCACATTAAGCAGGTA 2503
DB 2401 TTTCATGTTAGTGTAGCCAGCCACTGCCCTTTGTCCAAATGTGTCACATTAAGCAGGTA 2460
QY 2504 TCCCAGCATGAAATGTAATTTACTTTGGAAGTAACTTTGGAAAGTAACTTTGGAAAGAAATTCCTTCCTTTAAAT 2563
DB 2461 TCCCAGCATGAAATGTAATTTACTTTGGAAGTAACTTTGGAAAGTAACTTTGGAAAGAAATTCCTTCCTTTAAAT 2520
QY 2564 CAAAAACAAAAACAAAAACAAAAACACATTTCTAAATCTAGAGATAACTTTACTTTA 2623
DB 2521 CAAAAACAAAAACAAAAACAAAAACACATTTCTAAATCTAGAGATAACTTTACTTTA 2580
QY 2624 AATTCCTTCATTTTAGCAGTGATGATATGCAATGAGTGTGTAAAGCTTTGTAACTGGGGAAA 2683
DB 2581 AATTCCTTCATTTTAGCAGTGATGATATGCGTAAGTGCTGTAAAGCTTTGTAACTGGGGAAA 2640
QY 2684 TATTCACCTCATTAATGCCAGATTTACTGTATTTCCAAAAAGGCAATATTAAGGTAGA 2743
DB 2641 TATTCACCTCATTAATGCCAGATTTACTGTATTTCCAAAAAGGCAATATTAAGGTAGA 2700
QY 2744 TAGATGATTAGTATGATATTTCTACACATTTTGGNAATTAGAGAACATACAGAGGAA 2803
DB 2701 TAGATGATTAGTATGATATTTCTACACATTTTGGNAATTAGAGAACATACAGAGGAA 2760
QY 2804 TTTAGGGGCTTAAACATTTACGACTGAATGCACTTTAGTATAAAGGGCACAGTTTGTATAT 2863
DB 2761 TTTAGGGGCTTAAACATTTACGACTGAATGCACTTTAGTATAAAGGGCACAGTTTGTATAT 2820
QY 2864 TTTTAAATGAATACCAATTTTAAATTTTGTAGTATTTACCTGTTAAGAGATTTATAGTCTT 2923
DB 2821 TTTTAAATGAATACCAATTTTAAATTTTGTAGTATTTACCTGTTAAGAGATTTATAGTCTT 2880
QY 2924 TAAATTTTTAGGTTAAATTTTCTGCTGTGATATATATAGGAAATTTACTACTTTATGTC 2983
DB 2881 TAAATTTTTAGGTTAAATTTTCTGCTGTGATATATATAGGAAATTTACTACTTTATGTC 2940
QY 2984 CTGCTCTCTAAACTACATCTCGACTCGAGTCTGAGGATATATATCAACAGAGACACTTT 3043
DB 2941 CTGCTCTCTAAACTACATCTCGACTCGAGTCTGAGGATATATATCAACAGAGACACTTT 3000
QY 3044 TTGAGGCAATTTGAAAAACCAACCTACACTCTTCGGTCTTAGAGAGATCTGCTGCTCCC 3103
DB 3001 TTGAGGCAATTTGAAAAACCAACCTACACTCTTCGGTCTTAGAGAGATCTGCTGCTCCC 3060
QY 3104 AAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTTTCT 3163
DB 3061 AAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTTTCT 3120
QY 3164 GGTGATATCTGTGCTTCTCTAATATTCTGAAAGCTGCAATATTTTAGTAATACCTTCGGG 3223

Db 489 CTCAGGTCACCTCTCTGAAGCCCTCGTTACGGTGCATATCCCCACATCAAGAGCGGTGAG 548
Qy 584 GATCTTAAAGACCACTCAACAGAAAGTAAATAATGGAATTTGTCTAGGAGATCCAGG 643
Db 549 GATCTTAAAGACCACTCAACAGAAAGTAAATAATGGAATTTGTCTAGGAGATCCAGG 608
Qy 644 CATGAAGTAGAAAAATCAGAAAAATCAGTGAAGAACACAGATGCTTTGGGCAAAAATAGAGAA 703
Db 609 CATGAAGTAGAAAAATCAGAAAAATCAGTGAAGAACACAGATGCTTTGGGCAAAAATAGAGAA 668
Qy 704 TATTAATGTCGGGTGACAGCGCTTAAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACT 763
Db 669 TATAATGTCGGGTGACAGCGCTTAAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACT 728
Qy 764 AAGATTCTCGGGCCCAAGCCGAAGTCAAGTGGAGGAAGATCTCTGAAAAACGCTAT 823
Db 729 AAGATTCTCGGGCCCAAGCCGAAGTCAAGTGGAGGAAGATCTCTGAAAAACGCTAT 788
Qy 824 TCTCTAGATGACCTTGGAAATAGGCCAGGTCAGTTGTCTATCTTCTACATTTGACTCGGAG 883
Db 789 TCTCTAGATGACCTTGGAAATAGGCCAGGTCAGTTGTCTATCTTCTACATTTGACTCGGAG 848
Qy 884 AAAAAATGAGAGTAGAGCAAAATCTGGAACCTTCCAGGCTCTCAGAAACCTCTATAAAGGAT 943
Db 849 AAAAAATGAGAGTAGAGCAAAATCTGGAACCTTCCAGGCTCTCAGAAACCTCTATAAAGGAT 908
Qy 944 CGAATGSCCAAGTACAGGCGAGTGTGTCCAAACAAAGCGCTCAACCAACTATACAAAT 1003
Db 909 CGAATGSCCAAGTACAGGCGAGTGTGTCCAAACAAAGCGCTCAACCAACTATACAAAT 968
Qy 1004 GAGCTGAAGCCAGTGTGGCGGAAATCAAAATTCATAAAATGAGCAAAAGGAGAGATGTG 1063
Db 969 GAGCTGAAGCCAGTGTGGCGGAAATCAAAATTCATAAAATGAGCAAAAGGAGAGATGTG 1028
Qy 1064 CCCCCAGGTCCTGAGTCTGCATCACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAG 1123
Db 1029 CCCCCAGGTCCTGAGTCTGCATCACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAG 1088
Qy 1124 AATAGCTGCGAGTCGGTTCACCCCTGCGGAGATGACTCCC ---GTGACTCCCAGGTT 1180
Db 1089 AATAGCTGCGAGTCGGTTCACCCCTGCGGAGATGACTCCCAGGTCAGTCCCAGGTT 1148
Qy 1181 AAGAGTGAGGTTCAACAGCGCTGCCATCCCAAGCCACTAAGTCCAGATTTCCAGGCTCC 1240
Db 1149 AAGAGTGAGGTTCAACAGCGCTGCCATCCCAAGCCACTAAGTCCAGATTTCCAGGCTCC 1208
Qy 1241 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATCAAGAAAGTTTCAGGCACCTGCAAGAGAG 1300
Db 1209 AGTCTTTCTGAAAGTTCTCTCTCCCAAGCAATCAAGAAAGTTTCAGGCACCTGCAAGAGAG 1268
Qy 1301 ACCTGGTGGAAATGTCAGAAAGCAGTCTATPCCAATGGAGCGTCTCTTGGCCCAACAGCAG 1360
Db 1269 ACCTGGTGGAAATGTCAGAAAGCAGTCTATCCAATGGAGCGTCTCTTGGCCCAACAGCAG 1328
Qy 1361 GTGTTTTCACATCAGCTGCTTCCGTTGCTCTGCTATTGCAACCAAACTAGTCTAGAAACA 1420
Db 1329 GTGTTTTCACATCAGCTGCTTCCGTTGCTCTGCTATTGCAACCAAACTAGTCTAGAAACA 1388
Qy 1421 TATGCATCTTTTACATGGAAGAAATCTATTGTAGCCCTCACTTCAATCAACTCTTTTAAATCT 1480
Db 1389 TATGCATCTTTTACATGGAAGAAATCTATTGTAGCCCTCACTTCAATCAACTCTTTTAAATCT 1448
Qy 1481 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCAACCAAGGATCTATGGGCAAGCAAA 1540
Db 1449 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACCAACCAAGGATCTATGGGCAAGCAAA 1508
Qy 1541 AATGAAAAAGAGAGATTTTGGAGACACAGCCAGCTTGCAAAATGCAAGGAGACCCCT 1600
Db 1509 AATGAAAAAGAGAGATTTTGGAGACACAGCCAGCTTGCAAAATGCAAGGAGACCCCT 1568
Qy 1601 CACAGCCAGGGGTAGAAATGCCCTATTGTAGGTCGGGTGCTCCTGCTGCAAGATG 1660
Db 1569 CACAGCCAGGGGTAGAAATGCCCTATTGCTAAGGTCGGGTGCTCCTGCTGCAAGATG 1628

Qy 1661 GAAGCCAAAGCCCTCTCTCAGCAGGAGAGGAACACAGCCAGCTGAACCCAGAGCTG 1720
Db 1629 GAAGCCAGGCTCTCTCTCAGCAGGAGAGGAACACAGCCAGCTGAACCCAGAGCTG 1688
Qy 1721 AGGATCGCTCGCCACCCCTCACTGAACTTGGAACTTCAGGAAGTGCCTTGGAGGAAGG 1780
Db 1689 AGGATCGCTCGCCACCCCTCACTGAACTTGGAACTTCAGGAAGTGCCTTGGAGGAAGG 1748
Qy 1781 ATCAAAATGTCAAAAGCCCAAAATGGCTCTGAAGACGAAATCAGCAAGCCCAAGATTCCT 1840
Db 1749 ATCAAAATGTCAAAAGCCCAAAATGGCTCTGAAGACGAAATCAGCAAGCCCAAGATTCCT 1808
Qy 1841 GAGGATGTCGATCTAGATCTGAAGAGCTTAAGCATCTTCTTCACTGAAGAAAGAAAGC 1900
Db 1809 GAGGATGTCGATCTAGATCTGAAGAGCTTAAGCATCTTCTTCACTGAAGAAAGAAAGC 1868
Qy 1901 CCCCCATTCACGTAGAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 1960
Db 1869 CCCCCATTCACGTAGAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 1928
Qy 1961 TCCCCACCTATCAGGAAAGGCTGGAGCATGTGACAGCAGAGTGAAGAGTCTGTGGGTGGA 2020
Db 1929 TCCCCACCTATCAGGAAAGGCTGGAGCATGTGACAGCAGAAATGAAGATCTGTGGGTGGA 1988
Qy 2021 AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGGCTTTCTAAGAGAAATGGGAATGTG 2080
Db 1989 AGAGTTGCAGAAAGGAAACAAAGTGGAAATGCCAAGGCTTTCTAAGAGAAATGGGAATGTG 2048
Qy 2081 GGAAGAAACAACTGGCAAAACAAAGAAATCTAAGAGAGACAGGGAGAGAAAGTAAGGAA 2140
Db 2049 GGAAGAAACAACTGGCAAAACAAAGAAATTTAAGAGAGACAGGGAGAGAAAGTAAGGAA 2108
Qy 2141 GGTCTAGTGTGGAGATGGAGAAATGAGAAATCTGTAGAAAATGGTGACAGTCCGATCAA 2200
Db 2109 GGTCTAGTGTGGAGATGGAGAAATGAGAAATTTGTAGAAAATGGTGACAGTCCGATCAA 2169
Qy 2201 GATGATAACAGCTTCTCTCAAAACAAATCTCCCAAGAACCCCAAGTCTCTCAATTTGGTCG 2260
Db 2169 GATGATAACAGCTTCTCTCAAAACAAATTTCCCAAGAACCCCAAGTCTCTCAATTTGGTCG 2228
Qy 2261 AGTTTGTAGACACACCTTTGCTGAAGAAATCTACTACTCAGAAATCAGAAATCCAGGAT 2320
Db 2229 AGTTTGTAGACACACCTTTGCTGAAGAAATTTCACTACTCAGAAATCAGAAATCCAGGAT 2288
Qy 2321 GTGGAACCTCTGGGAGGAGAGTGGTCAAAGAGTCTCTGTGGAGAAACAGAAATAAGAGA 2380
Db 2289 GTGGAACCTTTGGGAGGAGAGTGGTCAAAGAGTCTCTGTGGAGAAACAGATAAAGAGA 2348
Qy 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGAAGAAATTCAGTCTCAGAAATCAGAAAT 2440
Db 2349 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGAAGAAATTCAGTCTCAGAAATCAGAAAT 2408
Qy 2441 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCCTTTGTCAAAATGTGATGCACATAAGCAG 2500
Db 2409 AAATTCATGTTAGTGTAGCGAGCCACTGCCCCCTTTGTCAAAATGTGATGCACATAAGCAG 2468
Qy 2501 GTATCCCGAGATGAATGTAAATTTACTTTGGAAGTAACTTTTGGAAAAAGAAATTCCTTCTTAA 2560
Db 2469 GTATCCCGAGATGAATGTAAATTTACTTTGGAAGTAACTTTTGGAAAAAGAAATTCCTTCTTAA 2528
Qy 2561 AATCAAAAAACAAAAACAAAAACAAAAACAAATCTAAATACTAGAGATAACTTTTAC 2620
Db 2529 AATCAAAAAACAAAAACAAAAACAAAAACAAATCTAAATACTAGAGATAACTTTTAC 2588
Qy 2621 TTAATTTCTTCATTTTAGCAGTGTATGATATGATGAAGTGTGTAAAGCTTGTACTCTGGG 2680
Db 2589 TTAATTTCTTCAT ----CAGTGTATGATATGATGAAGTGTGTAAAGCTTGTAACTGGG 2643
Qy 2681 AAATATCCACCTGATAAATAGCCAGATTTCTACTGTATTCCTCAAAAAAGCAATATTAAAGT 2740
Db 2644 AAATATCCACCTGATAAATAGCCAGATTTCTACTGTATTCCTCAAAAAAGCAATATTAAAGT 2703

2741 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAA 2800
|||||
2704 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTAGAGAACATACAGAA 2763
|||||
2801 GAAATTTAGGGGCTTAAACATTACGACTGAATGCAC-TTAGTATTAAGGGCCAGTTTGT 2860
|||||
2764 GAAATTTAGGGGCTTAAACATTACGACTGAATGCAC-TTAGTATTAAGGGCCAGTTTGT 2822
|||||
2861 TATTTTAAATGAATACCAATTTAAATTTTATTTTACCTGTTTAAAGAGATTAATTTAGT 2920
|||||
2823 TATTTTAAATGAATACCAATTTAAATTTTATTTTATTTTACCTGTTTAAAGAGATTAATTTAGT 2882
|||||
2921 CTTTAAATTTTATTTAGGTTAAATTTTCTGCTGATATATATGAGGAATTTACTACTTTAT 2980
|||||
2883 CTTTAAATTTTATTTAGGTTAAATTTTCTGCTGATATATATGAGGAATTTACTACTTTAT 2942
|||||
2981 GTCTGCTCTCTAAACTACATCTGAACTCGAGTCTGAGGTATTAATCAACAGAGCAC 3040
|||||
2943 GTCTGCTCTCTAAACTACATCTGAACTCGAGTCTGAGGTATA--ACAACAGAGCAC 3000
|||||
3041 TTTTGGAGCAATTTGAAACCAACCTACACTCTTGGTCTTTAGAGAGATCTGCTGTCT 3100
|||||
3001 TTTTGGAGCAATTTGAAACCAACCTACACTCTTGGTCTTTAGAGAGATCTGCTGTCT 3060
|||||
3101 CCCAAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTTT 3160
|||||
3061 CCCAAATAAGCTTTTGTATCTGCCAGTGAATTTACTGTACTCCAAATGATTTGCTTTCTTT 3120
|||||
3161 TCTGTGATATCTGTGCTTCTCATATTTACTGAAAGCTGCAATTTTAGTAATACCTTC 3220
|||||
3121 TCTGTGATATCTGTGCTTCTCATATTTACTGAAAGCTGCAATTTTAGTAATACCTTC 3180
|||||
3221 GGGATCACTGTCCTCCATCTTCCGTTGTAGAGCAAGTGAAGATTTTAAAGGAGGAGAA 3280
|||||
3181 GGGATCACTGTCCTCCATCTTCCGTTGTAGAGCAAGTGAAGATTTTAAAGGAGGAGAA 3240
|||||
3281 GAAAGACTGCTTTACACACTTTGAGCTCAGACCTCTAAACCCCTGATTTTCCCTTTATGAT 3340
|||||
3241 GAAAGACTGCTTTACACACTTTGAGCTCAGACCTCTAAACCCCTGATTTTCCCTTTATGAT 3300
|||||
3341 GTCCCTTTTGGAGACACTAATTTTAAATCTTACTAGCTCTGAATATATTGATTTT 3400
|||||
3301 GTCCCTTTTGGAGACACTAATTTTAAATCTTACTAGCTCTGAATATATTGATTTT 3360
|||||
3401 ATCAGATATCTCAGGTGAAATTTAAACCAACTATAGCCCTTTTCTTGGATGATTTT 3460
|||||
3361 ATCAGATATCTCAGGTGAAATTTAAACCAACTATAGCCCTTTTCTTGGATGATTTT 3420
|||||
3461 CTAGCTTTAAGGTTTGGGACATTTAAACTTGAGTACATTTTGTACAGTTGATAT 3520
|||||
3421 CTAGCTTTAAGGTTTGGGACATTTAAACTTGAGTACATTTTGTACAGTTGATAT 3480
|||||
3521 TCCAAATCTATGGATGGAGGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTGTACTGC 3580
|||||
3481 TCCAAATCTATGGATGGAGGAGAGGTGCTTAAAGCTGTAGGCTTTTCTTGTACTGC 3540
|||||
3581 ATTTATAGAGATTTAGCTTTAATATTTTATAGAGATTTAAACATCTGCTTTCTAGTC 3640
|||||
3541 ATTTATAGAGATTTAGCTTTAATATTTTATAGAGATTTAAACATCTGCTTTCTAGTC 3600
|||||
3641 TTACCTAGTCTGAACATTTTATTTCAATAAAGATTTTAAATTTAAATTTG 3690
|||||
3601 TTACCTAGTCTGAACATTTTATTTCAATAAAGATTTTAAATTTAAATTTG 3650
|||||

RESULT 5
LAF55696 standard; DNA; 3543 BP.
X D
X C
X AAF55696;
X
X 11-JUN-2001 (first entry)
X

DE Human EPLIN (epithelial protein lost in neoplasm)-alpha isoform DNA.
XX
KW Human; EPLIN; epithelial protein lost in neoplasm; EPLIN-alpha;
KW EPLIN-beta; tumour suppressor; tumour; cell proliferative disorder;
KW gene therapy; cancer; ss.
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT CDS 474..2276
FT /tag= a
FT /transl_except= "(pos: 825..827, aa: Val)"
FT /transl_except= "(pos: 1479..1481, aa: Asp)"
FT /transl_except= "(pos: 1627..1629, aa: Lys)"
FT /transl_except= "(pos: 1860..1861, aa: Ser)"
FT /transl_except= "(pos: 1968..1970, aa: Ser)"
FT /transl_except= "(pos: 2031..2033, aa: Ile)"
FT /transl_except= "(pos: 2088..2090, aa: Ser)"
FT /transl_except= "(pos: 2106..2108, aa: Ser)"
FT /product= "EPLIN (epithelial protein lost in neoplasm)"
XX
PN WO200118019-A1.
XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000WO-US24689.
XX
PR 08-SEP-1999; 99US-0153024.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Chang DB, Maul RS;
XX
DR WPI: 2001-244555/25.
DR P-PSDB; AAB67700.
XX
PT New tumor suppressor protein EPLIN, useful as a marker for diagnostic,
PT prognostic and therapeutic applications over the course of cell
PT proliferative disorders associated with EPLIN -
XX
PS Claim 4; Page 42-43; 59pp; English.
XX
CC The present sequence encodes a human EPLIN (epithelial protein lost in
CC neoplasm)-alpha isoform. The specification also describes EPLIN-beta.
CC EPLIN is a tumour suppressor protein, whose expression is altered in
CC multiple common human tumour types. EPLIN nucleic acids and proteins are
CC used in screening assays to detect molecules that specifically bind to
CC EPLIN nucleic acids, proteins or derivatives and thus have potential use
CC as agonist or antagonist of EPLIN, in particular molecules that affect
CC cell proliferation. Thus the assays are useful for screening molecules
CC with potential utility as anticancer drugs or lead compounds for drug
CC development. EPLIN nucleic acids, proteins are useful for detecting a
CC cell proliferative disorder in a subject. EPLIN polynucleotides are
CC useful in gene therapy techniques. EPLIN is useful as a marker that
CC can be diagnostically, prognostically and therapeutically used over
CC the course of a cell proliferative disorder associated with EPLIN.
XX
SQ Sequence 3543 BP; 1152 A; 711 C; 771 G; 909 T; 0 other;

Query Match 89.7%; Score 3324.4; DB 22; Length 3543;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 3371; Conservative 0; Mismatches 11; Indels 10; Gaps 3;
QY 302 AACACCGGAAATCTCTCCAGCACTTTAGAAAGGGGACCCCTGACTGTGTTAAAGAGAAG 361
|||
DB 159 AACACCGGAAATCTCTCCAGCACTTTAGAAAGGGGACCCCTGACTGTGTTAAAGAGAAG 218
|||
QY 362 TGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACT 421
|||
DB 219 TGGGAGAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACT 278
|||
QY 422 GAGATTAGGCACAGAGCAGACCACTCTCTGCTGTAAGTGACAGCCACGCTGCTTCTGGA 481
|||

Db 279 GAGATTAGGCACAGCAGACCATCTCTCTGCTGAAAGTGAAGCCACGCTGCTTCTGGA 338
Qy 482 GCCAAAGCTGACCAAGAAAGAAACAAATCCACCCAGATCTAGACTCAGGTCCACCTCTGTAA 541
Db 339 GCCAAAGCTGACCAAGAAAGAAACAAATCCACCCAGATCTAGACTCAGGTCCACCTCTGTAA 398
Qy 542 GCCCTGTTGAGGCTCGATATCCCAATCAAGACCGGTGAGGATCTTAAAGACCACTCA 601
Db 399 GCCCTGTTGAGGCTCGATATCCCAATCAAGACCGGTGAGGATCTTAAAGACCACTCA 458
Qy 602 ACAGAAAGTAAAGAAATGGAAGAAATGTTAGGAGATCCAGGATGAAGTAGAAAATCA 661
Db 459 ACAGAAAGTAAAGAAATGGAAGAAATGTTAGGAGATCCAGGATGAAGTAGAAAATCA 518
Qy 662 GAAATCAGTGAACACAGATGCTTCGGGCAAAATAGAGAAATATATATGTTCCGCTGAAC 721
Db 519 GAATCAGTGAACACAGATGCTTCGGGCAAAATAGAGAAATATATATGTTCCGCTGAAC 578
Qy 722 AGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACTAAGATTTCTCCGGGCCAA 781
Db 579 AGGCTTAAGATGATGTTTGAAGAAAGTGAACCAACTCAAACTAAGATTTCTCCGGGCCAA 638
Qy 782 AGCCGAAGTCAAGTGAAGAAAGATCTCTGAAGAAAGTGAACCAACTCAAACTAAGATTTCTCTAGATGACCTGGAA 841
Db 639 AGCCGAAGTCAAGTGAAGAAAGATCTCTGAAGAAAGTGAACCAACTCAAACTAAGATTTCTCTAGATGACCTGGAA 698
Qy 842 ATAGGCCACGTCAGTTGTCATCTTCTACATTTGACTCGGAGAAATAGAGTAGACGA 901
Db 699 ATAGGCCACGTCAGTTGTCATCTTCTACATTTGACTCGGAGAAATAGAGTAGACGA 758
Qy 902 AATCTGGAATTTCCAGCCTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAG 961
Db 759 AATCTGGAATTTCCAGCCTCTCAGAAACCTCTATAAGGATCGAATGGCCAAAGTACCAG 818
Qy 962 GCAGCTGTGTCCAAACAAAGCAGCTCAACCAATATACAAATGAGCTGAAAGCCAGTGT 1021
Db 819 GCAGCTGTGTCCAAACAAAGCAGCTCAACCAATATACAAATGAGCTGAAAGCCAGTGT 878
Qy 1022 GCGGAATCAAAATTCATAAAATGGAGCAAGAGAAATGTGCCCCAGGTCTCTGAGGTC 1081
Db 879 GCGGAATCAAAATTCATAAAATGGAGCAAGAGAAATGTGCCCCAGGTCTCTGAGGTC 938
Qy 1082 TGCAATCACCACATCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCTGGCAGTCCGT 1141
Db 939 TGCAATCACCACATCAGGAAGGGGAAAGATTTCTGCAAAATGAGAAATAGCTGGCAGTCCGT 998
Qy 1142 TCCACCCCTCCGAAGATGACTCCC---GTGACTCCAGGTTAAGAGTGAAGTCAACAG 1198
Db 999 TCCACCCCTCCGAAGATGACTCCCAGGTTAAGAGTGAAGTCAACAG 1058
Qy 1199 CTTGTCTATCCCAAGCCTAAGTCCAGATTCAGAGCTCCAGTCTTTCTGAAAAGTTCT 1258
Db 1059 CTTGTCTATCCCAAGCCTAAGTCCAGATTCAGAGCTCCAGTCTTTCTGAAAAGTTCT 1118
Qy 1259 CTTCCCAAGCAATGAAGATTTTCAAGCAGCTGGAAGAGAGACCTGCGTGAAGTGTGAG 1318
Db 1119 CTTCCCAAGCAATGAAGATTTTCAAGCAGCTGGAAGAGAGACCTGCGTGAAGTGTGAG 1178
Qy 1319 AAGACAGTCTATCCAAATGGAGGCTCTTTGGCCAAACCCAGCAGGTGTTTCAACATCAGCTGC 1378
Db 1179 AAGACAGTCTATCCAAATGGAGGCTCTTTGGCCAAACCCAGCAGGTGTTTCAACATCAGCTGC 1238
Qy 1379 TTCCGTTGCTCTATTTGCAACCAACAACTCAGTCTAGGAACATATGCATCTTTACATGGA 1438
Db 1239 TTCCGTTGCTCTATTTGCAACCAACAACTCAGTCTAGGAACATATGCATCTTTACATGGA 1298
Qy 1439 AGAATCTATTTGAGCCTCACTTCAATCAATCAATCTTTAAATCTAAGGGCAACTATGATGAA 1498
Db 1299 AGAATCTATTTGAGCCTCACTTCAATCAATCAATCTTTAAATCTAAGGGCAACTATGATGAA 1358
Qy 1499 GGCTTTGGGCACAGACCAACAGGATCTATGGCAAGCAAAATGAAACGAAGAGATT 1558
Db 1359 GGCTTTGGGCACAGACCAACAGGATCTATGGCAAGCAAAATGAAACGAAGAGATT 1418

Qy 1559 TTGGAGAGACCAGCCAGCTTGCNAATCAAGGGAGACCCCTTCACAGCCCGGGGTAGAA 1618
Db 1419 TTGGAGAGACCAGCCAGCTTGCNAATCAAGGGAGACCCCTTCACAGCCCGGGGTAGAA 1478
Qy 1619 GATGCCCTTATTTGCTAAGGTGGGTGCTCTGGCTGCAAGTATGAAAGCCCAAGCCCTCTCT 1678
Db 1479 AATGCCCTTATTTGCTAAGGTGGGTGCTCTGGCTGCAAGTATGAAAGCCCAAGCCCTCTCT 1538
Qy 1679 CAGCAGGAGAGGAGACCAAGCCAGCTGAACCAAGAGCTGAGGATCGCTGGCCACCC 1738
Db 1539 CAGCAGGAGAGGAGACCAAGCCAGCTGAACCAAGAGCTGAGGATCGCTGGCCACCC 1598
Qy 1739 CCCACTGAACCTTGAAGTTCAGGAAGTCCCTTGGAGGAGGGATCAAAATGTCAAAGGCC 1798
Db 1599 CCCACTGAACCTTGAAGTTCAGGAAGTCCCTTGGAGGAGGGATCAAAATGTCAAAGGCC 1658
Qy 1799 AAATGGCTCTCTGAAGACGAATCAGCAAGCCGAGTTCTCTGAAGGAAAGCCGCCCATTTCACTGTAGCA 1918
Db 1659 AAATGGCTCTCTGAAGACGAATCAGCAAGCCGAGTTCTCTGAAGGAAAGCCGCCCATTTCACTGTAGCA 1778
Qy 1859 CTGAAGAGCTTAAGACGATCTTCTTCAAGAGCCCAAAACTGTGTCCCCACCTATCAGGAAA 1978
Db 1719 CTGAAGAGCTTAAGACGATCTTCTTCAAGAGCCCAAAACTGTGTCCCCACCTATCAGGAAA 1838
Qy 1919 GCTTCATTTCAAAGCACTCTGTCAAGAGCCCAAAACTGTGTCCCCACCTATCAGGAAA 1978
Db 1779 GCTTCATTTCAAAGCACTCTGTCAAGAGCCCAAAACTGTGTCCCCACCTATCAGGAAA 1838
Qy 1979 GGCTGGAGCATGTGACAGCAGATGAAGTCTGTGGGTGGAAGAGTTGCGAAGAAAGGAAA 2038
Db 1839 GGCTGGAGCATGTGACAGCAGATGAAGTCTGTGGGTGGAAGAGTTGCGAAGAAAGGAAA 1898
Qy 2039 CAAAGTGGAAATGCCCAAGCTTCTAAGAGAAATGGAATGTGGGAAACCAACCTCGGCAA 2098
Db 1899 CAAAGTGGAAATGCCCAAGCTTCTAAGAGAAATGGAATGTGGGAAACCAACCTCGGCAA 1958
Qy 2099 AACAAAGAAATTAAGGAGACAGAGGAGAAAGTAAAGGAGGTCTAGTTTGGAGATG 2158
Db 1959 AACAAAGAAATTAAGGAGACAGAGGAGAAAGTAAAGGAGGTCTAGTTTGGAGATG 2018
Qy 2159 GAGAAATGAAATCTTTGTAGAAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCCTC 2218
Db 2019 GAGAAATGAAATCTTTGTAGAAAATGGTGCAGACTCCGATGAAGATGATAACAGCTTCCTC 2078
Qy 2219 AACAAAGAAATTAAGGAGACAGAGGAGAAAGTAAAGGAGGTCTAGTTTGGAGATG 2158
Db 2079 AACAAAGAAATTAAGGAGACAGAGGAGAAAGTAAAGGAGGTCTAGTTTGGAGATG 2018
Qy 2279 TTTGCTGAAGAAATTCACCTACTCAGAAATCAGAAATCCAGGATGGAACCTTGGAGGGA 2338
Db 2139 TTTGCTGAAGAAATTCACCTACTCAGAAATCAGAAATCCAGGATGGAACCTTGGAGGGA 2198
Qy 2339 GAAATGCTCAAAGAGCTCTCTGTGGAAGACAGATAAGAGAAATCGGTATTATGATGAG 2398
Db 2199 GAAATGCTCAAAGAGCTCTCTGTGGAAGACAGATAAGAGAAATCGGTATTATGATGAG 2258
Qy 2399 GATGAGATGAAGAGTGAAGAAATGCAATGATGCTGGCCCTTAAATTCATGTTAGTCTTA 2458
Db 2259 GATGAGATGAAGAGTGAAGAAATGCAATGATGCTGGCCCTTAAATTCATGTTAGTCTTA 2318
Qy 2459 GCGAGCCACTGCCCCCTTGTCAAAATGATGATGCTGGCCCTTAAATTCATGTTAGTCTTA 2518
Db 2319 GCGAGCCACTGCCCCCTTGTCAAAATGATGATGCTGGCCCTTAAATTCATGTTAGTCTTA 2378
Qy 2519 TAATTTACTTTGGAAGTAACTTTGGAAAGAAATTCCTTTTAAATCAAAACAAAAACAAA 2578
Db 2379 TAATTTACTTTGGAAGTAACTTTGGAAAGAAATTCCTTTTAAATCAAAACAAAAACAAA 2438
Qy 2579 AAAACAAAAACACATTTCTAAATCTAGAGATACTTTTAAATCTTTTAAATCTTTTAA 2638
Db 2439 AAAACAAAAACACATTTCTAAATCTAGAGATACTTTTAAATCTTTTAAATCTTTTAA 2493

RESULT 6	
AAI58169	
ID	AAI58169 standard; cDNA; 2905 BP.
XX	
AC	AAI58169;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 372.
XX	
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX	leukaemia; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	
WPI	2001-442253/47.
DR	P-PSDB; AAM39013.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
XX	such as central nervous system injuries .
XX	
PS	Claim 1; SEQ ID NO 372; 10078pp; English.
XX	
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC	the encoded polypeptides (AAM38642-AAW42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 2905 BP; 965 A; 617 C; 691 G; 632 T; 0 other;
Query Match	75.5%; Score 2797.6; DB 22; Length 2905;
Best Local Similarity	98.8%; Pred. No. 0;
Matches 2842; Conservative	0; Mismatches 24; Indels 10; Gaps 2
OV	4 CGCAGCAGCAGTAGGTGTAGCAGCTTGTCGCGACAGGTGCGCTAGTAGCGCCGG 63

4 CGACGAGCAGTAGGTGTACAGCTTGGTCGCGACAGGTGGCTAGGTAGAGCGCGGG 63

Db 40 CCGGGTCGACGATTTTCGTAGCAGCTTGGTCGACACAGGTGCGCTAGGTAGAGCGCGGG 99
QY 64 ACCTGTGACAGGCTGGTAGCAGCGCAGAGGAAAGGGGCTTTTAGCCAGGTAATTTTCAGT 123
Db 100 ACCTGTGACAGGCTGGTAGCAGCGCAGAGGAAAGGGGCTTTTAGCCAGGTAATTTTCAGT 159
QY 124 GTCTGTAGACAAGATGAATCATCTCCATTTAATAGACGCAATGACCTCCTCATTCATT 183
Db 160 GTCTGTAGACAAGATGAATCATCTCCATTTAATAGACGCAATGACCTCCTCATTCATT 219
QY 184 GAGGTAACAGCCAAAGAACTTCTCTTGTCAACAAGAACTCATCGCTATTTGTGA 243
Db 220 GAGGTAACAGCCAAAGAACTTCTCTTGTCAACAAGAACTCATCGCTATTTGTGA 279
QY 244 AATATTTCCAAATACAGAAAGAGCTGAAGAAACAAACATCGAGAAAGAGAAAGTAA 303
Db 280 AATATTTCCAAATACAGAAAGAGCTGAAGAAACAAACATCGAGAAAGAGAAAGTAA 339
QY 304 CACCGAAATCTCTCCAGCAGCTTTAGAAAGGGGACCTTGACTGTGTTTAAAGAAAGTG 363
Db 340 CACCGAAATCTCTCCAGCAGCTTTAGAAAGGGGACCTTGACTGTGTTTAAAGAAAGTG 399
QY 364 GGAGAACCCAGGGTGGAGCAGAGTCTCACACAGACTCTCTACGGAAACAGACGACTGA 423
Db 400 GGAGAACCCAGGGTGGAGCAGAGTCTCACACAGACTCTCTACGGAAACAGACGACTGA 459
QY 424 GATTAGSCACAGACAGACCATCCTCCTGCTGAAGTGACAAGCCAGCTGCTTTCTGGAG 483
Db 460 GATTAGSCACAGACAGACCATCCTCCTGCTGAAGTGACAAGCCAGCTGCTTTCTGGAG 519
QY 484 CAAAGCTGACCAAGAAACAAATCCACCCAGATCTAGACTCAGTCTACCTCCTGAAAG 543
Db 520 CAAAGCTGACCAAGAAACAAATCCACCCAGATCTAGACTCAGTCTACCTCCTGAAAG 579
QY 544 CCTCGTTTCAGGTCGATATCCACACATCAAGAGCGGTGAGGATCTTAAAGACCACTCAAC 603
Db 580 CCTCGTTTCAGGTCGATATCCACACATCAAGAGCGGTGAGGATCTTAAAGACCACTCAAC 639
QY 604 AGAAAGTAAAAAATGGAAATTTGCTAGGAGAAATCCAGGCATGAAGTAGAAAAATCAGA 663
Db 640 AGAAAGTAAAAAATGGAAATTTGCTAGGAGAAATCCAGGCATGAAGTAGAAAAATCAGA 699
QY 664 AATCAGTAAAAACACAGATGCTTCGGGCAAAAATAGAGAAATATAATGTTCCGCTGAACAG 723
Db 700 AATCAGTAAAAACACAGATGCTTCGGGCAAAAATAGAGAAATATAATGTTCCGCTGAACAG 759
QY 724 GCTTAAGATGATGTTGAGAAAGGTGAAACCACTCAAACTAAGATTTCTCGGGCCCAAG 783
Db 760 GCTTAAGATGATGTTGAGAAAGGTGAAACCACTCAAACTAAGATTTCTCGGGCCCAAG 819
QY 784 CCGAAGTCAAGTGGAAAGGAAGTCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAAAT 843
Db 820 CCGAAGTCAAGTGGAAAGGAAGTCTCTGAAAAACAGCTATTCTCTAGATGACCTGGAAAT 879
QY 844 AGGCCAGGTCAGTTGTCATCTTCTACATTTGACTCGGAGAAAAATGAGAGTAGACGAAA 903
Db 880 AGGCCAGGTCAGTTGTCATCTTCTACATTTGACTCGGAGAAAAATGAGAGTAGACGAAA 939
QY 904 TCTGGAACTTCCAGGCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCAGGC 963
Db 940 TCTGGAACTTCCAGGCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCAGGC 999
QY 964 AGCTGTGTCAAAACAAAGCAGCTCAACCACTATACAAATGAGCTCAAAAGCCAGTGGTG 1023
Db 1000 AGCTGTGTCAAAACAAAGCAGCTCAACCACTATACAAATGAGCTCAAAAGCCAGTGGTG 1059
QY 1024 CGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCCTGAGGTCTG 1083
Db 1060 CGAAATCAAAATTCATAAAATGGAGCAAAAGGAGAAATGTGCCCCAGGTCCTGAGGTCTG 1119
QY 1084 CATCAACCATACAGAAAGGGGAAAGATTTCTGCAAAATGAGAAATGAGCTTGGCAGTCCGTTT 1143
Db 1120 CATCAACCATACAGAAAGGGGAAAGATTTCTGCAAAATGAGAAATGAGCTTGGCAGTCCGTTT 1179

QY 1144 CACCCCTCCGAAAGATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCTGT 1203
Db 1180 CACCCCTCCGAAAGATGACTCCCGTGACTCCAGGTTAAGAGTGAGGTTCAACAGCCTGT 1239
QY 1204 CCATCCCAAGCCACTAAGTTCAGAGATTCCAGAGCTCCAGTCTTTCTGAAAGTCTCTCTCC 1263
Db 1240 CCATCCCAAGCCACTAAGTTCAGAGATTCCAGAGCTCCAGTCTTTCTGAAAGTCTCTCTCC 1299
QY 1264 CAAAGCAATGAAGAAAGTTTCAGGCACTTCGAGCACTTCGAGAGAGACCTGCGTGGAAATGTCAGAAGAC 1323
Db 1300 CAAAGCAATGAAGAAAGTTTCAGGCACTTCGAGCACTTCGAGAGAGACCTGCGTGGAAATGTCAGAAGAC 1359
QY 1324 AGTCTATCAATGGAGCGTCTCTTGGCCAAACAGCAGAGTGTTTTCACATCAGCTGCTCCG 1383
Db 1360 AGTCTATCAATGGAGCGTCTCTTGGCCAAACAGCAGAGTGTTTTCACATCAGCTGCTCCG 1419
QY 1384 TTGCTCTTATTGCAACACAACTCAGTCTAGGACATATGCATCTTTTACATGGAGAAAT 1443
Db 1420 TTGCTCTTATTGCAACACAACTCAGTCTAGGACATATGCATCTTTTACATGGAGAAAT 1479
QY 1444 CTATTGTAAAGCCTCACTTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGGCTT 1503
Db 1480 CTATTGTAAAGCCTCACTTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGGCTT 1539
QY 1504 TGGGCACAGACACACAAAGGATCTATGGGCAAGCAAAAATGAAAAACGAGAGATTTTGA 1563
Db 1540 TGGGCACAGACACACAAAGGATCTATGGGCAAGCAAAAATGAAAAACGAGAGATTTTGA 1599
QY 1564 GAGACCCAGCAGCTTGCAAAATGCAAGGGAGACCCCTCACAGCCCGAGGGGTAGAAGATGC 1623
Db 1600 GAGACCCAGCAGCTTGCAAAATGCAAGGGAGACCCCTCACAGCCCGAGGGGTAGAAGATGC 1659
QY 1624 CCCTATTCTTAAGTGGGTGCTTGGCTGCAAGTATGGAAGCCCAAGGCTCTCTCAGCA 1683
Db 1660 CCCTATTCTTAAGTGGGTGCTTGGCTGCAAGTATGGAAGCCCAAGGCTCTCTCAGCA 1719
QY 1684 GGAGAAAGAAACAAAGCCAGCTGAAACCAAGAGCTGAGGATCGCTCGCCACCCCCAC 1743
Db 1720 GGAGAAAGAAACAAAGCCAGCTGAAACCAAGAGCTGAGGATCGCTCGCCACCCCCAC 1779
QY 1744 TGAATCTTGAAGTTTCAGGAAGTGCTTGGAGAAAGGATCAAAATGTCAAAGCCCAATG 1803
Db 1780 TGAATCTTGAAGTTTCAGGAAGTGCTTGGAGAAAGGATCAAAATGTCAAAGCCCAATG 1839
QY 1804 GCCTCCTGAAGACGAAATCAGAGCCCGCAAGTTCCTGAGGATGTCGATCTAGATCTGAA 1863
Db 1840 GCCTCCTGAAGACGAAATCAGAGCCCGCAAGTTCCTGAGGATGTCGATCTAGATCTGAA 1899
QY 1864 GAAAGTGAAGACGATCTTCTTCACTGAAGAAAGAGCCGCCATTCAGTCTAGCAGCTTC 1923
Db 1900 GAAAGTGAAGACGATCTTCTTCACTGAAGAAAGAGCCGCCATTCAGTCTAGCAGCTTC 1959
QY 1924 ATTTCAAAGCACCTCTGTCAAGAGCCCAAAAATCTGTGCCCACTATCAGAAAGGCTG 1983
Db 1960 ATTTCAAAGCACCTCTGTCAAGAGCCCAAAAATCTGTGCCCACTATCAGAAAGGCTG 2019
QY 1984 GAGCATGTGACAGCAGAGTGAAAGTCTGTGGTGGAAAGTTCGAGAAAGGAAACAAATG 2043
Db 2020 GAGCATGTGACAGCAGAGTGAAAGTCTGTGGTGGAAAGTTCGAGAAAGGAAACAAATG 2079
QY 2044 GGAAATGCCAAGGCTTCTTAAGAGAAATGGAAATGTGGGAAAAACAACTGGCAAAACAA 2103
Db 2080 GGAAATGCCAAGGCTTCTTAAGAGAAATGGAAATGTGGGAAAAACAACTGGCAAAACAA 2139
QY 2104 AGAATCTAAAGAGAGACAGGAAAGAGAGTAAGGAAGTCTAGTTTGGAGATGGAGAA 2163
Db 2140 AGAATCTAAAGAGAGACAGGAAAGAGAGTAAGGAAGTCTAGTTTGGAGATGGAGAA 2199
QY 2164 TGAGAACTTTAGAAAAATGGTCAGACTCCGATGAAGATGATAACAGCTTCTCTCAACA 2223
Db 2200 TGAGAACTTTAGAAAAATGGTCAGACTCCGATGAAGATGATAACAGCTTCTCTCAACA 2259

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
P-PSDB; ABG08559.

English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes. Polymetase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques for restoring normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO

at ftp.wipo.int/pub/patent.html

sequence 3465 BP; 1112 A; 708 C; 1112 G;
length 3465;

```

75.3%; Score 2789.2; DB 23; Length 3789;
Query Match
Best Local Similarity 94.3%; Pred. No. 0; Gaps 33;
Matches 3266; Conservative 0; Mismatches 123; Indels 73;

```

[illegible]

XX Tang YT;
XX Tang YT;
XX Tang YT;

2224 ACAATCTCCAAAGAACCCCAAGTCTCTGAATTTGGTTCGAGTTTGTAGACACACCTTTGC 2283

2250 ACAATCTCCCAAGAACCCCAAGTCTCTGAATTTGGTTCGAGTTTGTAGACACACCTTTGC 2319

2280 TGAAGAATTCTACTACTCAGAAATCCCAAGATGTGGAACCTCTCGGAGGAGAAGT 2343

2284 TGAAGAATTCTACTACTCAGAAATCCCAAGATGTGGAACCTCTCGGAGGAGAAGT 2379

2320 TGAAGAATTCTACTACTCAGAAATCCCAAGATGTGGAACCTCTCGGAGGAGAAGT 2403

2344 GGTCAAAGAGCTCTCTGTGGAAGAACACAGATAAAGAGAAATCGGTATTATGATGAGGATCA 2439

2380 GGTCAAAGAGCTCTCTGTGGAAGAACACAGATAAAGAGAAATCGGTATTATGATGAGGATCA 2499

2404 GGATGAAGAGTGACAAATTTGCAATCATCTCGGGCCCTTAAATTCATGTTAGTGTAGCGAG 2463

2440 GGATGAAGAGTGACAAATTTGCAATCATCTCGGGCCCTTAAATTCATGTTAGTGTAGCGAG 2523

2464 CCACTGCCCTTTGTCAAATGTGATGCACATTAAGCAGGTATCCGAGCATGAAATGTAATT 2559

2500 CCACTGCCCTTTGTCAAATGTGATGCACATTAAGCAGGTATCCGAGCATGAAATGTAATT 2583

2524 TACTTGGAGGTAACTTTGGAAAAAGAAATTCCTCTTAAATCAAAAACAAAAAACA 2612

2560 TACTTGGAGGTAACTTTGGAAAAAGAAATTCCTCTTAAATCAAAAACAAAAAACA 2643

2584 ACAAAAAACACATTTCTAAATCTAGAGATAACTTTTACTTAAATTTCTTATTTAGCAGTG 2670

2613 --AAAAACACATTTCTAAATCTAGAGATAACTTTTACTTAAATTTCTTATTTAGCAGTG 2703

2644 ATGATATGATTAAGTCTGTAAAGCTTTGAACCTGGGGAATATTCACCTGTAATAGCC 2729

2671 ATGATATGATTAAGTCTGTAAAGCTTTGAACCTGGGGAATATTCACCTGTAATAG-C 2763

2704 CAGATTCTACTGTATTTCCAAAAGCAATATTAAGGTAGATAGATGATTAGTAGTATATT 2789

2730 CTTTTTTTCTTATTTCCATATGGCAATATTAAGGTAGACAGATGATTAGTAGTATATT 2823

2764 GTTACACACTATTTTGGAAATTAGAGAACATACAGAAGGAATTTTAGGGCTTTAAACATTAC 2849

2790 GTTACACACTATTTTGAATTAGAGATCATACAGAAGGATTTTAGGGCTTTAAACATTAC 2879

2824 GACTGAATGCACTTTGTATTAAGGGCAGAGTTTGTATATTTTAAATGAATACCA 2905

2850 GACTGAATGCACTTTGTATTAAGGGCAGAGTTTGTATATTTTAAATGAATACCA 2905

QY 842 ATAGCCAGGTCAGTTGTGATCTTCTACA-TTTGACTCGAGAAAATGAGAGTAGACG 900
DB ATAGCCAGGTCAGTTGTGATCTTCTACATTTTGACTCGGAGAAAATGAGAGTAGACG 603
QY 901 AAATCTGGAAGCTTCCAGCCCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCA 960
DB AAATCTGGAAGCTTCCAGCCCTCTCAGAAACCTCTATAAAGGATCGAATGGCCAAAGTACCA 663
QY 961 GGCAGCTGTGTCGAACAAAGCAGCTCAACCACTATACAAATGAGCTGAAAGCCAGTGG 1020
DB GGCAGCTGTGTCGAACAAAGCAGCTCAACCACTATACAAATGAGCTGAAATGCCAGCGG 723
QY 1021 TGGCGAAATCAAAATTCATATAAATGGAGCAAAAGAGAGAAATGTGCCCCAGTCTGAGGT 1080
DB TGGCGAAATCAAAATTCATATAAATGGAGCAAAAGAGAGAAATGTGCCCCAGTCTGAGGT 783
QY 1081 CTGCATACCCA-TCAGAGAGGGGAAAAGATTTCTGCAAAATGAGAAATGAGCTGGCAGTCC 1139
DB CTGCATACCCCAATTCAGAGAGGGGAAAAGATTTCTGCAAAATGAGAAATGAGCTGGCAGTCC 843
QY 1140 GTTCCACCCCTGCCGAAGATGACTCCC---GTGACTCCCAGGTTAAGAGTAGAGTTCAAC 1196
DB GTTCCACCCCTGCCGAAGATGACTCCCAGGTTGACTCCCAGGTTAAGAGTAGAGTTCAAC 903
QY 1197 AGCCTGTCCATCCCAGGCACCTAAGTCCAGATTCAGAGCCCTCAGTCTTTCTGAAAAGTT 1256
DB AGCCTGTCCATCCCAGGCACCTAAGTCCAGATTCAGAGCCCTCAGTCTTTCTGAAAAGTT 963
QY 1257 CTCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTC 1316
DB CTCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCGTGGAAATGTC 1023
QY 1317 AGAAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAGGTTTTCACATCAGCT 1376
DB AGAAGACAGTCTATCCAATGGAGCGTCTCTTGGCCAAACAGCAGGTTTTCACATCAGCT 1083
QY 1377 GCTTCGGTTGCTCCTATTGCAACAAC-AAACTCAGTCTAGGAACATATGCATCTTTACAT 1435
DB GCTTCGGTTGCTCCTATTGCAACAACAAAACTCAGTCTAGGAACATATGCATCTTTACAT 1143
QY 1436 GGAAGAACTTATGTGAAGCCTCACTTCAATCAACTCTTTTAAATCT--AAGGCCACTATG 1493
DB GGAAGAACTTATGTGAAGCCTCACTTCAATCAACTCTTTTAAATCTTAAGGGCAAACTATG 1203
QY 1494 ATGAAGCCTTTGGGCACAGACCACACAGGATCT--ATGGGCAAGCAAAAATGAAAAC-G 1550
DB ATGAAGCCTTTGGGCACAGACCACACAGGATCTTTATTTGGGCAAGCAAAAATGAAAACGG 1263
QY 1551 AAGAGATTTTGGAGAGACCAG-CCGAGCTTGCAAAATGCAAGGGAGACCCCTCACAGCCCA 1609
DB AAGAGATTTTGGAGAGACCAG-CCGAGCTTGCAAAATGCAAGGGAGACCCCTCACAGCCCA 1323
QY 1610 GGGGTAGAAGATGCCCTTATGCTAAGTGGGTGCTCGCTGCTGCAAGTATGGAAGCCCAAG 1669
DB GGGGTAGAAGATGCCCTTATGCTAAGTGGGTGCTCGCTGCTGCAAGTATGGAAGCCCAAG 1383
QY 1670 GCCTCCTCTCAGCAGAGAGAGAGACAGCCAGCTGAAACCAAGAGCTGAGGATCGCC 1729
DB GCCTCCTCTCAGCAGAGAGAGAGACAGCCAGCTGAAACCAAGAGCTGAGGATCGCC 1443
QY 1730 TGGCCACCCCACTCAACTTGGAACTTCAGGAAGTGGCTTCGAGGAAGGGATCAAAATG 1789
DB TGGCCACCCCACTCAACTTGGAACTTCAGGAAGTGGCTTCGAGGAAGGGATCAAAATG 1503
QY 1790 TCAAGGCCCAAAATGGCCTCCTGAAGACGAAATCAGCAAGCCCGAAAGTTCTCTGAGGATGTC 1849
DB TCAAGGCCCAAAATGGCCTCCTGAAGACGAAATCAGCAAGCCCGAAAGTTCTCTGAGGATGTC 1563
QY 1850 GATCTAGATCTGAAGAGCTTAACAGATCTTCTTCACTGAAGAAAGAGAGCCGCCCAATTC 1909
DB GATCTAGATCTGAAGAGCTTAACAGATCTTCTTCACTGAAGAAAGAGAGCCGCCCAATTC 1623

QY 1910 ACTGTAGCAGCTTCAATTTCAAAGCAGCTCTGTCAAGAGCCCAAAAACCTGTGTCCCAACC- 1968
DB ACTGTAGCAGCTTCAATTTCAAAGCAGCTCTGTCAAGAGCCCAAAAACCTGTGTCCCAACC 1683
QY 1969 TATCAGGAAAGGCTGGAGCATGTCAAGCAGAGTGAAGAGTCTGTGGTGGAGAGATTGC 2028
DB TATCAGGAAAGGCTGGAGCATGTCAAGCAGAGTGAAGAGTCTGTGGTGGAGAGATTGC 1743
QY 2029 AGAAAGGAACAAGTGGAAAATGCCAAGCTTCTTAAGAGAAATGGGAATGTGGGAAAAAC 2088
DB AGAAAGGAACAAGTGGAAAATGCCAAGCTTCTTAAGAGAAATGGGAATGTGGGAAAAAC 1803
QY 2089 AACCTGGCAAAAACAAAGAAATCTAAAGGAGAGACAGGGAAGAGAAAGTAAGG--AAGGTCTAT 2146
DB AACCTGGCAAAAACAAAGAAATCTTAAGGAGAGACAGGGAAGAGAAAGTAAGGAAAGGTCCAT 1863
QY 2147 AGTTTGAGATGAGAAATGAGAAATCTT-----GTAGAAAATGGTGCAGACTCCCGATTGAAG 2201
DB AGTTTGAGATGAGAAATGAGAAATCTTGTAGAAAATGGTGCAGACTCCCGATTGAAG 1923
QY 2202 ATGATAACAGCTTCTTCAAACAACAA-----TCTCCAAAGAACCCCAAGTCTTCTTGAAT 2252
DB ATGATAACAGCTTCTTCAAACAACAACTTCTCCAAAGAACCCCAAGTCTTCTTGAAT 1983
QY 2253 ATTGGTCGAGTTTGTAGACAACACC---TTTGTGGAAGAAATTCACACT-CAGAATCAG 2308
DB ATTGGTCGAGTTTGTAGACAACACCCTTTTGTGGAAGAAATTCACACTCAGAAATCAG 2043
QY 2309 AAATCCAGGATGTGGAACTCTGGGAGGAGAAAGTGGTCAAAGAGTCTCTCTGGGAAGAA 2368
DB AAATCCAGGATGTGGAACTCTGGGAGGAGAAAGTGGTCAAAGAGTCTCTCTGGGAAGAA 2103
QY 2369 CAGATAAGAGAAATCGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTTGCAATG 2428
DB CAGATAAGAGAAATCGTATTATGATGAGGATGAGGATGAAGAGTGACAAATTTGCAATG 2163
QY 2429 ATGCT-GGGCTTTAAATTCATGTTAGTGTAGCGAGCCTGCGCTTTGTCAAAATGTGA 2487
DB ATGCTGGGCTTTAAATTCATGTTAGTGTAGCGAGCCTGCGCTTTGTCAAAATGTGA 2223
QY 2488 TGCACATAACGAGTATCCGAGGATGAAATGTAATTTACTTTGGAAGTAACCTTTGGAAAAAG 2547
DB TGCACATAACGAGTATCCGAGGATGAAATGTAATTTACTTTGGAAGTAACCTTTGGAAAAAG 2283
QY 2548 AATTCCTTCTTAAATCAAAAAACAAAAACAAAAACAAAAA--ACACATTTCTTAAATAC 2605
DB AATTCCTTCTTAAATCAAAAAACAAAAACAAAAACAAAAACAAAAATCTTAAATAC 2343
QY 2606 TAGAGATAACTTTTACTTAAATTTCTTCAATTTTAGCAGTGATGATATGATAGTCTGTAA 2665
DB TAGAGATAACTTTTACTTAAATTTCTTCAATTTTAGCAGTGATGATATGATAGTCTGTAA 2403
QY 2666 GCCTTTGAAGTGGGAAATATCCA-CCTGATAATAGCCAGATTTCTACTGTATTTCCAA 2724
DB GCCTTTGAAGTGGGAAATATTCACCCCTGATAATAGCCAGATTTCTACTGTATTTCCAA 2463
QY 2725 AAGCAATATTAAGGTAGATA-GATGATTAG-TAGTATTTTGTACACACTATTTTGGAA 2782
DB AAGCAATATTAAGGTAGGAGGATGATTAGTATATTTGTACACACTATTTTGGAA 2523
QY 2783 TTAGAGAACATACAGAGGAAATTTAGGGGCTTTAAACATTAACGACT-GAATGCACTTTAGT 2841
DB TTAGAGAACATACAGAGGAAATTTAGGGGCTTTAAACATTAACGACTGGAATGCACTTTAGT 2583
QY 2842 ATAAAGGGCAGTTTGTATTAATTTT---AAATGAATACCAATTTAATTTTGTAGTATTT 2898
DB ATAAAGGGCAGTTTGGTATATTTTAAATTTTAAATTTTAAATTTTAAATTTTGTAGTATTT 2643
QY 2899 ACCTGTTAAGAGATTTTAGTC-TTTAAATTTTGTAGTAAATTTTGTCTGTGTAT 2957
DB ACCTGTTAAGAGATTTTAGTCTTTTAAATTTTAAATTTTGTAGTAAATTTTGTCTGTGTAT 2703
QY 2958 ATATGAGGAATTTTACTACTTTTATGTCTGTCTCT-TAAACTACATCCT-GAACTCGACGT 3015

1361 GTGTTTACATCAGCTGCTCCGTTTGCTCCTATTGCAACACAACTCAGTCTAGGAACA 1420
Db |||||
420 GTGTTTACATCAGCTGCTCCGTTTGCTCCTATTGCAACACAACTCAGTCTAGGAACA 479
Qy |||||
1421 TATGCATCTTTACATGGAAGAAATCTATTGTAAAGCCTCACCTCAATCAACTCTTTAAATCT 1480
Db |||||
480 TATGCATCTTTACATGGAAGAAATCTATTGTAAAGCCTCACCTCAATCAACTCTTTAAATCT 539
Qy |||||
1481 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACAAAGAGATCTATGGCGAAGCAAA 1540
Db |||||
540 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACACAAAGAGATCTATGGCGAAGCAAA 599
Qy |||||
1541 AATGAACACGAAGAGATTTTGGAGAGACCACCCAGCTTGCAAATGCAAGGAGACCCCT 1600
Db |||||
600 AATGAACACGAAGAGATTTTGGAGAGACCACCCAGCTTGCAAATGCAAGGAGACCCCT 659
Qy |||||
1601 CACAGCCAGGGGTAGAGATGCCCCCTATTGCTAAAGTGGGTGCTCGGCTGCAAGTATG 1660
Db |||||
660 CACAGCCAGGGGTAGAGATGCCCCCTATTGCTAAAGTGGGTGCTCGGCTGCAAGTATG 719
Qy |||||
1661 GAAGCCAAAGGCTCTCTCAGCAGGAGAAAGACCAAGCCAGCTGAACCCAGAGAGCTG 1720
Db |||||
720 GAAGCCAAAGGCTCTCTCAGCAGGAGAAAGACCAAGCCAGCTGAACCCAGAGAGCTG 779
Qy |||||
1721 AGGATCCGCTGGCCACCCCTCCTGAACCTTGAAGTTTCAGGAAGTGCCTTGGAGGAAGG 1780
Db |||||
780 AGGATCCGCTGGCCACCCCTCCTGAACCTTGAAGTTTCAGGAAGTGCCTTGGAGGAAGG 839
Qy |||||
1781 ATCAAAATGTCAAAAGCCCAATGGCCCTCTGAAGACGAAATCAGCAAGCCCGAAGTCTCT 1840
Db |||||
840 ATCAAAATGTCAAAAGCCCAATGGCCCTCTGAAGACGAAATCAGCAAGCCCGAAGTCTCT 899
Qy |||||
1841 GAGGATGCGATCTGATCTGAAGAACTAAGACGATCTTCTTCACTGAAGAAAGAAAGC 1900
Db |||||
900 GAGGATGCGATCTGATCTGAAGAACTAAGACGATCTTCTTCACTGAAGAAAGAAAGC 959
Qy |||||
1901 CGCCCATCTAGTAGCAGCTTCAATTTCAAGCACTCTCTCAAGAGCCCAAAACTGTG 1960
Db |||||
960 CGCCCATCTAGTAGCAGCTTCAATTTCAAGCACTCTCTCAAGAGCCCAAAACTGTG 1019
Qy |||||
1961 TCCCACTATCAGGAAGGCTGGAGCATCTCAGACAGAGTCAAGAGTCTGTGGTGGA 2020
Db |||||
1020 TCCCACTATCAGGAAGGCTGGAGCATCTCAGACAGAGTCAAGAGTCTGTGGTGGA 1079
Qy |||||
2021 AGAGTTGCAGAAAGGAACAAAGTGGAAATGCCAAAGCTTCTAAGAAAGATGGGAATGTG 2080
Db |||||
1080 AGAGTTGCAGAAAGGAACAAAGTGGAAATGCCAAAGCTTCTAAGAAAGATGGGAATGTG 1139
Qy |||||
2081 GGAAGAAACAACTGGCAAAACAAAGATCTAAAGGAGACAGGGAGAGAGTAAGGAA 2140
Db |||||
1140 GGAAGAAACAACTGGCAAAACAAAGATCTAAAGGAGACAGGGAGAGAGTAAGGAA 1199
Qy |||||
2141 GGTATAGTTTGGAGATGGAGATGGAATCTTGTAGAAAATGGTGCAGACTCCCGATGAA 2200
Db |||||
1200 GGTATAGTTTGGAGATGGAGATGGAATCTTGTAGAAAATGGTGCAGACTCCCGATGAA 1259
Qy |||||
2201 GATGATAACAGCTTCTTCAAAACAACTCTCCACAAAGACCCCAAGTCTCTGAATTTGGTCG 2260
Db |||||
1260 GATGATAACAGCTTCTTCAAAACAACTCTCCACAAAGACCCCAAGTCTCTGAATTTGGTCG 1319
Qy |||||
2261 AGTTTGTAGAACAACTTCTTGTGAAGAAATTCACCTACTCAGAAATCCCAAGAT 2320
Db |||||
1320 AGTTTGTAGAACAACTTCTTGTGAAGAAATTCACCTACTCAGAAATCCCAAGAT 1379
Qy |||||
2321 GTGGAATCTGGGAGGAGAGTGTCAAGAGCTCTGTGGAAGAAACAGATAAGAGA 2380
Db |||||
1380 GTGGAATCTGGGAGGAGAGTGTCAAGAGCTCTGTGGAAGAAACAGATAAGAGA 1439
Qy |||||
2381 AATCGGTATTATGATCAGGATGAGGATGAAGAGTGCATATTCATGATGCTGGGCCCT 2440
Db |||||
1440 AATCGGTATTATGATCAGGATGAGGATGAAGAGTGCATATTCATGATGCTGGGCCCT 1499

2441 AAATTCATGTTAGTGTAGGAGCCACTGCCCTTTGTCAAAATGTGATGCACATAAGCAG 2500
Db |||||
1500 AAATTCATGTTAGTGTAGGAGCCACTGCCCTTTGTCAAAATGTGATGCACATAAGCAG 1559
Qy |||||
2501 GTATCCAGCATGAATGTAATTTACTTGGAGTACCTTTGGAAA*GAATTCCTCTTAA 2560
Db |||||
1560 GTATCCAGCATGAATGTAATTTACTTGGAGTAACTTTGGAAAAGAATTCCTCTTAA 1619
Qy |||||
2561 AATCAAAAACAAAACAAAACAAAACAAAACAACTTCTAAATACTAGAGATAAATTTAC 2620
Db |||||
1620 AATCAAAAACAAAACAAAACAAAACAAAACAACTTCTAAATACTAGAGATAAATTTAC 1679
Qy |||||
2621 TTAATTCCTTCAATTTTACAGTATGATGATGATGCTGTAAAGCTTTGAAGCTGGG 2680
Db |||||
1680 TTAATTCCTTCAATTTTACAGTATGATGATGCTGTAAAGCTTTGAAGCTGGG 1739
Qy |||||
2681 AATATTCACCTGATTAATAGCCAGATCTACTGTATTTCCAAAAGGCAATATTAAAGT 2740
Db |||||
1740 AATATTCACCTGATTAATAGCCAGATCTACTGTATTTCCAAAAGGCAATATTAAAGT 1799
Qy |||||
2741 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTTAGAACATACAGAG 2800
Db |||||
1800 AGATAGATGATTAGTAGTATATTGTTACACACTATTTTGGAAATTTAGAACATACAGAG 1859
Qy |||||
2801 GAATTTAGGGGCTTAAACATTAAGACTGAATGCACCTTTAGTATAAAGGCCACAGTTGTA 2860
Db |||||
1860 GAATTTAGGGGCTTAAACATTAAGACTGAATGCACCTTTAGTATAAAGGCCACAGTTGTA 1919
Qy |||||
2861 TATTTTAAATGAATACCAATTAATTTTGTAGTATTTTACCTGTTAAGAGATATTATAGT 2920
Db |||||
1920 TATTTTAAATGAATACCAATTAATTTTGTAGTATTTTACCTGTTAAGAGATATTATAGT 1979
Qy |||||
2921 CTTTAAATTTTGTAGTATTTTCTGCTGTGATATATAGGAATTTTACTACTTTAT 2980
Db |||||
1980 CTTTAAATTTTGTAGTATTTTCTGCTGTGATATATAGGAATTTTACTACTTTAT 2039
Qy |||||
2981 GTCCTGCTCTCTAAACTACATCCTGAACCTGACGCTCTGAGGTATAATACACAGAGCAC 3040
Db |||||
2040 GTCCTGCTCTCTAAACTACATCCTGAACCTGACGCTCTGAGGTATAATACACAGAGCAC 2099
Qy |||||
3041 TTTTGGAGCAATTGAAAACCAACCTACACTCTTCCGGTGTCTAGAGAGATCTGCTGCT 3100
Db |||||
2100 TTTTGGAGCAATTGAAAACCAACCTACACTCTTCCGGTGTCTAGAGAGATCTGCTGCT 2159
Qy |||||
3101 CCCAAATTAAGCTTTTGTATCTGCAGTGAATTTACTGTACTCCAAATGATTCCTTTCT 3160
Db |||||
2160 CCCAAATTAAGCTTTTGTATCTGCAGTGAATTTACTGTACTCCAAATGATTCCTTTCT 2219
Qy |||||
3161 TCTGTGATATCTGTGCTTCTCATAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTC 3220
Db |||||
2220 TCTGTGATATCTGTGCTTCTCATAATTTACTGAAAGCTGCAATATTTTAGTAATACCTTC 2279
Qy |||||
3221 GGGATCACTGTCCCCCATCTTCCGTGTAGAGCAAGTGAAGAGTTTAAAGGAGGAAGAA 3280
Db |||||
2280 GGGATCACTGTCCCCCATCTTCCGTGTAGAGCAAGTGAAGAGTTTAAAGGAGGAAGAA 2339
Qy |||||
3281 GAAAGAACTGTCTTACACCACTTGAGCTCAGACCTCTAAACCCCTGTATTTCCCTTATGAT 3340
Db |||||
2340 GAAAGAACTGTCTTACACCACTTGAGCTCAGACCTCTAAACCCCTGTATTTCCCTTATGAT 2399
Qy |||||
3341 GTCCCTTTTGGAGACACTAAATTTTAAATCTTACTAGCTCTGAAATATATGATTTTT 3400
Db |||||
2400 GTCCCTTTTGGAGACACTAAATTTTAAATCTTACTAGCTCTGAAATATATGATTTTT 2459
Qy |||||
3401 ATCAGATATTCTCAGGGTGAATTTAAACCAACTATAGGCTTTTCTTGGGATGATTTT 3460
Db |||||
2460 ATCAGATATTCTCAGGGTGAATTTAAACCAACTATAGGCTTTTCTTGGGATGATTTT 2519
Qy |||||
3461 CTAGCTTAAAGGTTTGGGACATTTAAACCTTGGAGTACATTTGTTGACACAGTTGATAT 3520
Db |||||
2520 CTAGCTTAAAGGTTTGGGACATTTAAACCTTGGAGTACATTTGTTGACACAGTTGATAT 2579
Qy |||||
3521 TCCAAATTTGATGATGGAGGGAGAGGTGCTTTAAGCTGTAGGCTTTTCTTGTACTGC 3580

2580 TCCAAATTGATGGAGGAGGTTCTTAAGCTGAGGCTTTCTTTGTACTGC 2639
3581 ATTTATAGAGATTAGCTTTAATTTTATAGAGATGTAAGCAATTCGTCTTTCTAGTC 3640
2640 ATTTATAGAGATTAGCTTTAATTTTATAGAGATGTAAGCAATTCGTCTTTCTAGTC 2699
3641 TTACTAGTCTGAACATTTTATTCATTAAGAGATTTTAAATTAATAATTG 3690
2700 TTACCTAGTCTGAACATTTTATTCATTAAGAGATTTTAAATTAATAATTG 2749

LT 9
3566
AAZ33566 standard; cDNA; 2783 BP.

AAZ33566;
08-DEC-1999 (first entry)
Human breast tumour-associated EST 26.
Expressed sequence tag; EST; human; breast; cancer; cytostatic;
medicaments; gene therapy; treatment; fat metabolism; ss.
Homo sapiens.
DE19813835-A1.
23-SEP-1999. 98DE-1013835.
20-MAR-1998; 98DE-1013835.
20-MAR-1998; 98DE-1013835.
(META-) METAGEN GES GENOMFORSCHUNG MBH.
Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
WPI: 1999-528979/45.
P-PSDB; AAY48487.
Human nucleic acid sequences and protein products from normal breast
tissue, useful for breast cancer therapy
Claim 1a; 113-114; 206pp; German.
This invention describes novel human nucleic acid sequences from normal
breast tissue which have cytotostatic activity. The nucleic acid sequences
can be used to produce and isolate full-length gene sequences. They can
be used to express proteins, which can be used as tools to find an
activity against breast cancer. The sequences can be used in sense or
antisense form. They are especially useful for medicaments for gene
therapy to treat breast cancer and for treating illnesses associated
with fat metabolism. AAZ33541-233610 represent expressed sequence tags
described in the method of the invention.

Sequence 2783 BP; 901 A; 527 C; 608 G; 747 T; 0 other;
Query Match 72.8%; Score 2696.6; DB 20; Length 2783;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2699; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
1001 AATGAGCTGAAGAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGGCAAT 1060
40 AATGAGCTGAAGAGCCAGTGGTGGCGAAATCAAAATTCATAAAATGGAGCAAAAGGCAAT 99
1061 GTGCCCCCAGGCTCTGAGGTCTGCAATCACCCTCAGGAAGGGGAAAGATTCTCGCAAT 1120
100 GTGCCCCCAGGCTCTGAGGTCTGCAATCACCCTCAGGAAGGGGAAAGATTCTCGCAAT 159
1121 GAGATAGCTGGCAGTCCGTTCCACCTCCGCGAAGATGACTCCCGTGACTCCAGGTT 1180

Db 160 GAGAAATAGCTGGCAGTCCGTTCCACCCCTGCCAAGATGATCCCGTGACTCCAGGTT 219
QY 1181 AAGAGTGAAGTTCAACAGCGCTGTCATCCCAAGCCACTAAGTCAGATTCCAGAGCCCTCC 1240
Db 220 AAGAGTGAAGTTCAACAGCGCTGTCATCCCAAGCCACTAAGTCAGATTCCAGAGCCCTCC 279
QY 1241 AGTCTTTCTGAAGTTCTCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 1300
Db 280 AGTCTTTCTGAAGTTCTCCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 339
QY 1301 ACCTGGTGGMATGTCAGAAAGCAGTCTATCCAAATGGAGCGTCTCTTGGCCACACAGCAG 1360
Db 340 ACCTGGTGGMATGTCAGAAAGCAGTCTATCCAAATGGAGCGTCTCTTGGCCACACAGCAG 399
QY 1361 GTGTTTCACATCAGCTGCTTCCGTTGCTCTATTGCAACAACAACTCAGTCTAGGAACA 1420
Db 400 GTGTTTCACATCAGCTGCTTCCGTTGCTCTATTGCAACAACAACTCAGTCTAGGAACA 459
QY 1421 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 1480
Db 460 TATGCATCTTTACATGGAAGAATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCT 519
QY 1481 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAA 1540
Db 520 AAGGCCAACTATGATGAAGGCTTTGGGCACAGACCAACAAGGATCTATGGGCAAGCAAA 579
QY 1541 AATGAAAACGAAGAGATTTTGGAGAGACCCAGCTTGCAAATGGAAGGAGACCCCT 1600
Db 580 AATGAAAACGAAGAGATTTTGGAGAGACCCAGCTTGCAAATGGAAGGAGACCCCT 639
QY 1601 CACAGCCCAAGGCTAGAGATGCCCCCTATTGCTAAGGTGGGTGCTCTGGCTGCAAGTATG 1660
Db 640 CACAGCCCAAGGCTAGAGATGCCCCCTATTGCTAAGGTGGGTGCTCTGGCTGCAAGTATG 699
QY 1661 GAAGCCCAAGGCTCTCTCAGCAGGAGGAAGGAGCAAGCCAGCTGAAACCAAGAGCTG 1720
Db 700 GAAGCCCAAGGCTCTCTCAGCAGGAGGAAGGAGCAAGCCAGCTGAAACCAAGAGCTG 759
QY 1721 AGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTTGAAGTTCAGGAAGTCCCTTGGAGGAGGG 1780
Db 760 AGGATCGCTGGCCACCCCTCACTGAACTTGAAGTTCAGGAAGTCCCTTGGAGGAGGG 819
QY 1781 ATCAAAATGTCAAAGCCCAAAATGGCTCTCTGAAGAGCAATCAGCAAGCCGAGTTCTCT 1840
Db 820 ATCAAAATGTCAAAGCCCAAAATGGCTCTCTGAAGAGCAATCAGCAAGCCGAGTTCTCT 879
QY 1841 GAGGATCTCAGTCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAAAGAGC 1900
Db 880 GAGGATCTCAGTCTAGATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAAAGAGC 939
QY 1901 CGCCCATTCAGTCTAGCAGCTTCAATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 1960
Db 940 CGCCCATTCAGTCTAGCAGCTTCAATTTCAAAGCACCTCTGTCAAGAGCCCAAAACTGTG 999
QY 1961 TCCCCACCTATCAGGAAAGGCTGGAGCATGTCTGAGCAGAGTGAAGAGTCTGTGGGTGGA 2020
Db 1000 TCCCCACCTATCAGGAAAGGCTGGAGCATGTCTGAGCAGAGTGAAGAGTCTGTGGGTGGA 1059
QY 2021 AGAGTTGCAAGAAAGAAACAAGTGGAAATGCCAAGCTTCTTAAGAGCAATGGGAATGTG 2080
Db 1060 AGAGTTGCAAGAAAGAAACAAGTGGAAATGCCAAGCTTCTTAAGAGCAATGGGAATGTG 1119
QY 2081 GGAAGAAACCACTGCAAGAAACAAGTGGAAATGCCAAGCTTCTTAAGAGCAATGGGAATGTG 2140
Db 1120 GGAAGAAACCACTGCAAGAAACAAGTGGAAATGCCAAGCTTCTTAAGAGCAATGGGAATGTG 1179
QY 2141 GGTTCATAGTTTGGAGATGGAGATGAGAAATCTTGTAGAAAATGGTGCAGACTCCGATGAA 2200
Db 1180 GGTTCATAGTTTGGAGATGGAGATGAGAAATCTTGTAGAAAATGGTGCAGACTCCGATGAA 1239
QY 2201 GATGATAACAGCTTCTCTCAACAAACAATCTCCACAGAGACCCCAAGTCTCTGAATTTGGTCG 2260
Db 1240 GATGATAACAGCTTCTCTCAACAAACAATCTCCACAGAGACCCCAAGTCTCTGAATTTGGTCG 1299

The polynucleotide sequences given in AAC59449 to AAC59497 encode the human secreted proteins given in AAB34092 to AAB34140. AAB34141 to AAB34216 represent human secreted polypeptide sequences and proteins homologous to them, which are given in the exemplification of the present invention. Human secreted proteins have activities based on the tissue and cells the genes are expressed in. Examples of activities include: antiarthritic; immunosuppressive; antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective; ophthalmological; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases, hyperproliferative disorders e.g. neoplasms or cancer of the breast or liver, cardiovascular disorders, infections caused by disorders, angiogenesis, nervous system disorders. The polypeptides can bacteria, viruses and fungi and ocular disorders. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used in the exemplification of the present invention.

Sequence 2267 BP; 741 A; 416 C; 490 G; 617 T; 3 other;

Query Match	59.6%;	Score 2208.8;	DB 21;	Length 2267;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 2207;	Conservative	Mismatches 0;	Indels 0;	Gaps 0;
1496	GAAGGCTTTGGGCACAGACACACAGGATCTATGGCAGCAAAATGAAACGAAAG	1555		
32	GAAGGCTTTGGGCACAGACACACAGGATCTATGGCAGCAAAATGAAACGAGAG	91		
1556	ATTTTGAGAGACAGCCAGCTTGCAATGCAAGGAGACCCCTCACAGCCAGGGGTA	1615		
92	ATTTTGAGAGACAGCCAGCTTGCAATGCAAGGAGACCCCTCACAGCCAGGGGTA	151		
1616	GAAGATCCCTATTGTAAGGTGGGTGCTCTGGCTGCAAGTATGGAAGCCAGGCTCC	1675		
152	GAAGATCCCTATTGTAAGGTGGGTGCTCTGGCTGCAAGTATGGAAGCCAGGCTCC	211		
1676	TCTCAGCAGGAGAGAGACAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGCCA	1735		
212	TCTCAGCAGGAGAGAGACAGCCAGCTGAAACCAAGAGCTGAGGATCGCTGCCA	271		
1736	CCCCCCTGAACTTTGGAAGTTCAGGAAGTGCCTTGGAGGAGGAGATCAAAATGTCAAAG	1795		
272	CCCCCCTGAACTTTGGAAGTTCAGGAAGTGCCTTGGAGGAGGAGATCAAAATGTCAAAG	331		
1796	CCCAATGGCCTTCTGAGACCGAATCAGCAGCCCGAAGTTCCTGAGGATGTGATCTA	391		
332	CCCAATGGCCTTCTGAGACCGAATCAGCAGCCCGAAGTTCCTGAGGATGTGATCTA	1915		
1856	GATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAGAGCCGCCCTTCACTGTA	451		
392	GATCTGAAGAGCTAAGACGATCTTCTTCACTGAAGGAGAGCCGCCCTTCACTGTA	1975		
1916	GCAGCTTCATTTCAAGACCTCTGTCAAGAGCCCAAAATGTTGCCCACTATCAGG	511		
452	GCAGCTTCATTTCAAGACCTCTGTCAAGAGCCCAAAATGTTGCCCACTATCAGG	2035		
1976	AAAGGCTGGAGCATGTTCAGAGCAGGTGAAGAGTCTGTGGGTGAAGAGTTGCAGAAAGG	571		
512	AAAGGCTGGAGCATGTTCAGAGCAGGTGAAGAGTCTGTGGGTGAAGAGTTGCAGAAAGG	2095		
2036	AAACAAGTGGAATGCAAGGCTTCTAAGAAGATGGGAATGGGAATAAACAAGCTGG	631		
572	AAACAAGTGGAATGCAAGGCTTCTAAGAAGATGGGAATGGGAATAAACAAGCTGG	2155		
2096	CAAAACAAAGAAATCTAAAGGAGAGACAGGAGAGAGTAGGAGGCTCATAGTTGGAG			

632	CAAAACAAAGAAATCTAAAGAGAGACAGGAGAGAAAGTAAGGAGGTCTATAGTTGGAG	691		
2156	ATGAGAAATGAGAAATCTTTAGAAAATGGTGCAGATCCGATGAAGATGATAACAGCTTC	2215		
692	ATGAGAAATGAGAAATCTTTAGAAAATGGTGCAGATCCGATGAAGATGATAACAGCTTC	751		
2216	CTCAAAACAAATCTCCACAGAACCCCAAGTCTCTGAATTTGGTTCGAGTTTGTAGACAAC	2275		
752	CTCAAAACAAATCTCCACAGAACCCCAAGTCTCTGAATTTGGTTCGAGTTTGTAGACAAC	811		
2276	ACCTTTGCTGAAGATTCACCTACTCAGAAATCCAGAAATCCAGGATGTGGAACTCTGGAG	2335		
812	ACCTTTGCTGAAGATTCACCTACTCAGAAATCCAGAAATCCAGGATGTGGAACTCTGGAG	871		
2336	GGAGAAATGGTCAAGAGCTCTCTGTGGAAGAACAGATTAAGAGAAATCGGTATTATCAT	2395		
872	GGAGAAATGGTCAAGAGCTCTCTGTGGAAGAACAGATTAAGAGAAATCGGTATTATCAT	931		
2396	GAGGATGAGGATGAAGAGTGCACAAATTCGAATGATGCTGGGCTTTAAATTCATGTAGTG	2455		
932	GAGGATGAGGATGAAGAGTGCACAAATTCGAATGATGCTGGGCTTTAAATTCATGTAGTG	991		
2456	TTAGCGAGCAGCTGCCCTTTGTCAAAATGTGATGCACATTAAGCAGGTATCCAGCATGAA	1051		
992	TTAGCGAGCAGCTGCCCTTTGTCAAAATGTGATGCACATTAAGCAGGTATCCAGCATGAA	1111		
2516	ATGTAATTTACTTTGGAAATCTTTGGAAAAGAAATTCCTTTTAAATCAAAACAAAAC	1171		
1052	ATGTAATTTACTTTGGAAATCTTTGGAAAAGAAATTCCTTTTAAATCAAAACAAAAC	1231		
2576	AAAAAACACAAAAAACACATCTTAATCTAGAGATTAATCTTAAATTTCTTCAATTT	1291		
1112	AAAAAACACAAAAAACACATCTTAATCTAGAGATTAATCTTAAATTTCTTCAATTT	1351		
2636	TAGCAGTGATGATGATAGTCTGTAAAGCTTTGTAAGCTGGGAAATATTCACCTGA	1411		
1172	TAGCAGTGATGATGATAGTCTGTAAAGCTTTGTAAGCTGGGAAATATTCACCTGA	1471		
2696	TAAAGCCAGATCTCTGATTTCCCAAAAGGCAATTAATTAAGGTAGATGATGATAGT	1531		
1232	TAAAGCCAGATCTCTGATTTCCCAAAAGGCAATTAATTAAGGTAGATGATGATAGT	1591		
2756	AGTATATTGTTACACACTATTTTGAATTTAGAGACATACAGAGGAAATTTAGGGGCTTA	1651		
1292	AGTATATTGTTACACACTATTTTGAATTTAGAGACATACAGAGGAAATTTAGGGGCTTA	1711		
2816	AACATTTACGACTGAATGCACTTTAGTATAAGGGGACAGTTTGTATTTTAAATGAAT	1771		
1352	AACATTTACGACTGAATGCACTTTAGTATAAGGGGACAGTTTGTATTTTAAATGAAT	1831		
2876	ACCAATTTAAATTTTGTATTTTACCTGTTAAGAGATTTTAAATTTTAAATTTTAA	1891		
1412	ACCAATTTAAATTTTGTATTTTACCTGTTAAGAGATTTTAAATTTTAAATTTTAA	1951		
2936	GTTAATTTTCTGCTGTGATATATAGAGAAATTTACTATTTATGCTGCTCTCTAA	2011		
1472	GTTAATTTTCTGCTGTGATATATAGAGAAATTTACTATTTATGCTGCTCTCTAA	2071		
2996	CTACATCTGAACCTGACGCTCGAGGTATTAATACACAGAGCACTTTTGGAGCAATTG	2131		
1532	CTACATCTGAACCTGACGCTCGAGGTATTAATACACAGAGCACTTTTGGAGCAATTG	2191		
3056	AAAAACCAACCTACCTCTCGGTGCTTAGAGAGATCTGCTGCTCTCCCAATAAGCTTTT	2251		
1592	AAAAACCAACCTACCTCTCGGTGCTTAGAGAGATCTGCTGCTCTCCCAATAAGCTTTT	2311		
3116	GTATCTGCAGGATTTTACTGTACTCCAAATGATGCTTTCTTTTGTGATATCTGT	2371		
1652	GTATCTGCAGGATTTTACTGTACTCCAAATGATGCTTTCTTTTGTGATATCTGT	2431		
3176	GCTTCTCTAATTTACTGAAAGCTGCAATTTTGTAGTAATCTCTCGGATCAGCTGCC	2491		

1712 GCTTCTCATAAATTACTGAAAGCTGCAATATTTTAGTAATACCTTCGGGATCACTGTCCCC 1771
3236 CATCTTCGCGTGTAGAGCAAAAGTGAAGATTTTAAAGGAGGAAGAAAGAACTGTCTTA 3295
1772 CATCTTCGCGTGTAGAGCAAAAGTGAAGATTTTAAAGGAGGAAGAAAGAACTGTCTTA 1831
3296 CACCACCTTGAGCTCAGACCTCTAAACCCCTGTATTTCCCTTATGATGTCCCTTTTGGAG 3355
1832 CACCACCTTGAGCTCAGACCTCTAAACCCCTGTATTTCCCTTATGATGTCCCTTTTGGAG 1891
3356 CACTAAATTTTAAATCTACTAGCTCTGAAATATATTTGATTTTATCAAGATTTCTCA 3415
1892 CACTAAATTTTAAATCTACTAGCTCTGAAATATATTTGATTTTATCAAGATTTCTCA 1951
3416 GGGTGAATTAACCAACTATAGCCCTTTTCTTGGGATGATTTTCTAGTCTTAAGTTT 3475
1952 GGGTGAATTAACCAACTATAGCCCTTTTCTTGGGATGATTTTCTAGTCTTAAGTTT 2011
3476 GGGACATTATAAATCTGAGTACATTTGTTGACACAGTTGATTTCCAAATTTGATGGA 3535
2012 GGGACATTATAAATCTGAGTACATTTGTTGACACAGTTGATTTCCAAATTTGATGGA 2071
3536 TGGGAGGGAGAGGTGCTTTAAAGCTGTAGGCTTTTCTTTGACTGCAATTTATAGAGATTTA 3595
2072 TGGGAGGGAGAGGTGCTTTAAAGCTGTAGGCTTTTCTTTGACTGCAATTTATAGAGATTTA 2131
3596 GCTTTAATATTTTTPAGAGATGTAAAAATTTCTGCTTTTCTTTAGTCTTACCTAGTCTGAAA 3655
2132 GCTTTAATATTTTTPAGAGATGTAAAAATTTCTGCTTTTCTTTAGTCTTACCTAGTCTGAAA 2191
3656 CATTTTATTCATTAAGATTTTAAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 3705
2192 CATTTTATTCATTAAGATTTTAAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 2241

RESULT 11
AAC93483
ID AAC93483 standard; cDNA; 2158 BP.
XX AAC93483;
XX
XX
XX 16-FEB-2001 (first entry)
XX Human secreted protein gene 5 SEQ ID NO:15.
XX
XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytosolic; cardiac; vasotropic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; vulnery; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorders; cancer; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.
XX
XX Homo sapiens.
XX
XX WO200061626-A1.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09066.
XX
XX 09-APR-1999; 99US-0128698.
XX 20-JAN-2000; 2000US-0176926.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-619227/59.
XX P-PSDB; AAB51831.
XX
XX New nucleic acid molecules encoding 49 human secreted proteins for

diagnosing, preventing or ameliorating medical conditions and used for food additives or preservatives -
Claim 1; Page 435; 516pp; English.
CC Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 - AAB51927 represent alternative polypeptides encoded by the genes, and amino acid sequences with which they share homology. The genes and proteins have activities dependent on the tissues and cells in which they are expressed. Examples of their activities include immunosuppressive; antiarthritic; antirheumatic; antiproliferative; cytosolic; cardiac; vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; and vulnerary. The secreted proteins, polynucleotides, antagonists and agonists may be useful in treating, preventing and/or diagnosing diseases and disorders such as autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular ischaemia, e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrates, vitamins, minerals, cofactors and other nutritional components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are used in the isolation and characterisation of the proteins and polynucleotides of the invention.

Sequence 2158 BP; 714 A; 383 C; 458 G; 603 T; 0 other;
Query Match 56.2%; Score 2083.6; DB 21; Length 2158;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 2138; Conservative 0; Mismatches 14; Indels 6; Gaps 4;
QY 1553 GAGATTTTGGAGAGACCCAGCTTGCATAATGCAAGGAGACCCCTCACAGCCAGGG 1612
DB 1 GAGATTTTGGAGAGACCCAGCTTGCATAATGCAAGGAGACCCCTCACAGCCAGGG 60
QY 1613 GTAGAAGATGCCCTATTGTCTAAGGTGGTGTCTGTGCTCAAGTATGGAAGCAAGGCC 1672
DB 61 GTAGAAGATGCCCTATTGTCTAAGGTGGTGTCTGTGCTCAAGTATGGAAGCAAGGCC 120
QY 1673 TCCTCTCAGCAGGAGAGAGCAAGCAAGCTGTAACCAAGAGCTGAGGATCGCTGG 1732
DB 121 TCCTCTCAGCAGGAGAGAGCAAGCAAGCTGTAACCAAGAGCTGAGGATCGCTGG 180
QY 1733 CCACCCCTTCACTGAACTTGGAGTTCAGGAGTCCCTTGGAGGAGGATCAAAATGTCA 1792
DB 181 CCACCCCTTCACTGAACTTGGAGTTCAGGAGTCCCTTGGAGGAGGATCAAAATGTCA 240
QY 1793 AAGCCCAATGGCTCTCTGAAGAGCAATCAGCAAGCCGCAAGTTCCTGAGGATGTGAT 1852
DB 241 AAGCCCAATGGCTCTCTGAAGAGCAATCAGCAAGCCGCAAGTTCCTGAGGATGTGAT 300
QY 1853 CTAGATCTGAAGAGCTAAGACGATCTTTCTTCTGTAAGGAGGAGGAGCCGCCCTTCACT 1912
DB 301 CTAGATCTGAAGAGCTAAGACGATCTTTCTTCTGTAAGGAGGAGGAGCCGCCCTTCACT 360
QY 1913 GTAGCAGCTTCAATTTCAAGACCTCTGTCAAGAGCCCAAAACTGTGTCCCGACCTATC 1972
DB 361 GTAGCAGCTTCAATTTCAAGACCTCTGTCAAGAGCCCAAAACTGTGTCCCGACCTATC 420
QY 1973 AGGAAGGCTGGAGCATGTCTAGAGCAGAGTGAAGAGTCTGTGGTGGAGAGTGTGAGAA 2032
DB 421 AGGAAGGCTGGAGCATGTCTAGAGCAGAGTGAAGAGTCTGTGGTGGAGAGTGTGAGAA 480
QY 2033 AGGAAGCAAGTGGAAATGCCCAAGGCTTCTTAAAGAGAAATGGGAATGTGGGAAAAACAAC 2092
DB 481 AGGAAGCAAGTGGAAATGCCCAAGGCTTCTTAAAGAGAAATGGGAATGTGGGAAAAACAAC 540

DR P-PSDB; ABB56420.
XX Human protein able to suppress growth of cancer cells and its coding
PT sequence -
PS
XX Claim 5; Page 21-22 Disclosure; 37pp; Chinese.
XX
CC The invention relates to novel human proteins (ABB56417-ABB56425) with
CC cancer suppressing function, the encoding polynucleotides
CC (AB19970-AB19978), the process for preparing the polypeptide, the
CC application of the polypeptide in treating diseases such as cancer, the
CC antagonist of the polypeptide and its medical function and the
CC application of the polynucleotide.
XX
XX Sequence 1754 BP; 586 A; 377 C; 426 G; 361 T; 4 other;
SQ
Query Match 46.5%; Score 1721.4; DB 24; Length 1754;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1747; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
829 AGATGACCTGGAAATAGGCCCGAGTGTGATCTTCTACATTTGACTCGGAGAAAA 888
D 1 AGATGACCTGGAAATAGGCCCGAGTGTGATCTTCTACATTTGACTCGGAGAAAA 60
Q 889 TGAGAGTAGACGAATCTGGAATCTCCAGCCTCTCAGAACCTCTATAAAGGATCGAAT 948
D 61 TGAGAGTAGACGAATCTGGAATCTCCAGCCTCTCAGAACCTCTATAAAGGATCGAAT 120
Q 949 GGCCAACTACAGGAGCTGTGTCCTCAAAACAAAGAGAGCTCAACCAACTATACAAATGAGCT 1008
D 121 GGCCAACTACAGGAGCTGTGTCCTCAAAACAAAGAGAGCTCAACCAACTATACAAATGAGCT 180
Q 1009 GAAAGCAGTGTGGGGAATCAAAATTCATAAATGGAGCAAAAGAGAAATGTGCCCCC 1068
D 181 GAAAGCAGTGTGGGGAATCAAAATTCATAAATGGAGCAAAAGAGAAATGTGCCCCC 240
Q 1069 AGTCTGAGGTGTGATCACCCTCAGAGAGGGGAAAGATTTCTGCAATGAGAAATAG 1128
D 241 AGTCTGAGGTGTGATCACCCTCAGAGAGGGGAAAGATTTCTGCAATGAGAAATAG 300
Q 1129 CTTGGCAGTCCGTTCCACCTCCGGAAGATGACTCCGTTGACTCCAGGTTAAGAGTGA 1188
D 301 CTTGGCAGTCCGTTCCACCTCCGGAAGATGACTCCGTTGACTCCAGGTTAAGAGTGA 360
Q 1189 GGTTCAACAGCTGTCCATCCCAAGGCACTAAGTCCAGAT - TCCAGAGCCTCCAGTCTTT 1247
D 361 GGTTCAACAGCTGTCCATCCCAAGGCACTAAGTCCAGATCTCCAGAGCTCCAGTCTTT 420
Q 1248 CTGAAAGTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCG 1307
D 421 CTGAAAGTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAGACCTGCG 480
Q 1308 TGGAAATGTCAGAGACAGTCTATCCAGAGCGTCTCTGGCCACCAGCAGGTGTTTC 1367
D 481 TGGAAATGTCAGAGACAGTCTATCCAGAGCGTCTCTGGCCACCAGCAGGTGTTTC 540
Q 1368 ACATCAGCTGCTCCGTTGCTCTATTCGAAACAACTCAGTCTAGGAACATATGCAT 1427
D 541 ACATCAGCTGCTTCGTTGCTCTATTCGAAACAACTCAGTCTAGGAACATATGCAT 600
Q 1428 CTTTACATGGAAGAACTATTGTAAGCCTCACTCAATCAACTCTTTAAATCTAAGGGCA 1487
D 601 CTTTACATGGAAGAACTATTGTAAGCCTCACTCAATCAACTCTTTAAATCTAAGGGCA 660
Q 1488 ACTATGATGAGGCTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAAAAATGAAA 1547
D 661 ACTATGATGAGGCTTTGGGCACAGACCACACAGGATCTATGGGCAAGCAAAAATGAAA 720
Q 1548 ACGAAGAGATTTTGGAGAGACCCAGCCAGCTTCAAAATCAAGGAGAGCCCTCACAGCC 1607
D 721 ACGAAGAGATTTTGGAGAGACCCAGCCAGCTTCAAAATCAAGGAGAGCCCTCACAGCC 780
Q 1608 CAGGGGTAGAAGATGCCCCCTATTGTTCAAGGTGGGTGTCTCGTGGCTGCAAGTATGGAAGCCA 1667

781 CAGGGGTAGAAGATGCCCTATTGCTAAGGTGGGTGCTCTGGCTGCAAGTATGGAAGCCA 840
Q 1668 AGGCCTCTCTCAGCAGAGAGGAAGACAAGCCAGCTGAAACCAAGAAAGTGTAGGATCG 1727
D 841 AGGCCTCTCTCAGCAGAGAGGAAGACAAGCCAGCTGAAACCAAGAAAGTGTAGGATCG 900
Q 1728 CTTGCCACCCCTCCACTGAACTTGGAACTTCAGGAAGTGCCTTGGAGGAAGGGATCAAAA 1787
D 901 CTTGCCACCCCTCCACTGAACTTGGAACTTCAGGAAGTGCCTTGGAGGAAGGGATCAAAA 960
Q 1788 TGTCAAAAGCCCAATGGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGTTCCTGAGGATG 1847
D 961 TGTCAAAAGCCCAATGGGCTCTCTGAAGACGAAATCAGCAAGCCCAAGTTCCTGAGGATG 1020
Q 1848 TCGATCTAGATCTGAAGAACTAAGACGATCTTCTTCACTGAAGGAAAGAAAGCCGCCAT 1907
D 1021 TCGATCTAGATCTGAAGAACTAAGACGATCTTCTTCACTGAAGGAAAGAAAGCCGCCAT 1080
Q 1908 TCACCTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAAACCTGTGTCCCCAC 1967
D 1081 TCACCTGTAGCAGCTTCATTTCAAAGCACCTCTGTCAAGAGCCCAAAAACCTGTGTCCCCAC 1140
Q 1968 CTATCAGGAAGAGGCTGGAGCATGTCCAGAGCAGTGTGAAGAGTCTGTGGGTGGAGAGTTG 2027
D 1141 CTATCAGGAAGAGGCTGGAGCATGTCCAGAGCAGTGTGAAGAGTCTGTGGGTGGAGAGTTG 1200
Q 2028 CAGAAAGGAAACAACTGGGAAATGCAAGGCTTCTTAAAGAAATGGGAAATGTGGGAAAAA 2087
D 1201 CAGAAAGGAAACAACTGGGAAATGCAAGGCTTCTTAAAGAAATGGGAAATGTGGGAAAAA 1260
Q 2088 CAACCTGCAAAACAAAGAAATCTTAAAGGAGAGACAGGGAAGAGAGTAAAGAAAGTCTATA 2147
D 1261 CAACCTGCAAAACAAAGAAATCTTAAAGGAGAG - CAGGGAAGAGAGTAAAGAAAGTCTATA 1319
Q 2148 GTTTCGAGATGAGAAATGAGAAATCTTGTAGAAAATGGTGCAGATCCGATGAAGATGATA 2207
D 1320 GTTTCGAGATGAGAAATGAGAAATCTTGTAGAAAATGGTGCAGATCCGATGAAGATGATA 1379
Q 2208 ACAGCTTCTCTCAAAACAACATCTCCACAGAACCAAGTCTCTGAAATGGTGCAGTTTG 2267
D 1380 ACAGCTTCTCTCAAAACAACATCTCCACAGAACCAAGTCTCTGAAATGGTGCAGTTTG 1439
Q 2268 TAGACAACACCTTTCTCTGAAAGAAATCTACTCTCAGAAATCAGAAATCCCGAGATGTGGAAC 2327
D 1440 TAGACAACACCTTTCTCTGAAAGAAATCTACTCTCAGAAATCAGAAATCCCGAGATGTGGAAC 1499
Q 2328 TCTGGAGGGGAGAGTGTCTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGT 2387
D 1500 TCTGGAGGGGAGAGTGTCTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCGGT 1559
Q 2388 ATTATGATGAGATGAGATGAAGAGTCAAAAATTCGAAATGATGCTGGGCTTAAATTTCA 2447
D 1560 ATTATGATGAGATGAGATGAAGAGTCAAAAATTCGAAATGATGCTGGGCTTAAATTTCA 1619
Q 2448 TCTTATGTTAGCAGGACCTGCCCCCTTGTCAAAATGTATGCAATAGCAGGATATCCC 2507
D 1620 TCTTATGTTAGCAGGACCTGCCCCCTTGTCAAAATGTATGCAATAGCAGGATATCCC 1679
Q 2508 AGCATGAAATGTAAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTCTTAAAAATCAA 2567
D 1680 AGCATGAAATGTAAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTCTTAAAAATCAA 1739
Q 2568 AACAAAAACAAAAA 2582
D 1740 AAAAAAAAAAAAAA 1754

RESULT 13
AAI59955
ID AAI59955 standard; cDNA; 1713 BP.
XX
AC AAI59955;

XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 3944.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX Homo sapiens.
XX WO200153312-A1.
XX PN
XX PD
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 200WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX P-PSDB; AAM40799.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 3944; 100789p; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 1713 BP; 549 A; 401 C; 417 G; 346 T; 0 other;
XX
XX Query Match
XX Best Local Similarity 43.0%; Score 1595; DB 22; Length 1713;
XX Matches 1642; Conservative 0; Mismatches 60; Indels 1; Gaps 1;
XX
XX 22 TAGCAGCTTGTCGCGACAGTGGCGCTAGGTAGAGCGCGGACCTGTGACAGGGCTGGT 81
XX 6 TAGCAGCTTGTCGCGACAGTGGCGCTAGGTAGAGCGCGGACCTGTGACAGGGCTGGT 65
XX
XX 82 AGCAGCCGAGAGGAAGCGCGCTTTAGCCAGGTATTTTCAGTGTCTGTAGACAAGATGGA 141
XX 66 AGCAGCCGAGAGGAAGCGCGCTTTAGCCAGGTATTTTCAGTGTCTGTAGACAAGATGGA 125

QY 142 ATCATCTCCATTTAATAGACGGCAATGGACCTCACTATCATTCATTGAGGGTAACAGCCAAAGA 201
DB 126 ATCATCTCCATTTAATAGACGGCAATGGACCTCACTATCATTCATTGAGGGTAACAGCCAAAGA 185
QY 202 ACTTTCTCTGTCAACAAGAACAAAGTCATCGGCTATTGTGGAATATTCTTCAAGTACCA 261
DB 186 ACTTTCTCTGTCAACAAGAACAAAGTCATCGGCTATTGTGGAATATTCTTCAAGTACCA 245
QY 262 GAAAGCAGCTGAAGAAACAAACATGAGAGGAGAGAGTAAACACGCAAAATCTCTCCCA 321
DB 246 GAAAGCAGCTGAAGAAACAAACATGAGAGGAGAGAGTAAACACGCAAAATCTCTCCCA 305
QY 322 GCACTTTAGAAAGGGGACCTCTGACTGTGTTAAAGAAAGAGTGGGAAACCCAGGGCTGGG 381
DB 306 GCACTTTAGAAAGGGGACCTCTGACTGTGTTAAAGAAAGAGTGGGAAACCCAGGGCTGGG 365
QY 382 AGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACTGAGATTAGGCACAGAGCAGA 441
DB 366 AGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACTGAGATTAGGCACAGAGCAGA 425
QY 442 CCATCTCTCTGTGAAGTGCACAAAGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAAGA 501
DB 426 CCATCTCTCTGTGAAGTGCACAAAGCCACGCTGCTTCTGGAGCCAAAGCTGACCAAGAAGA 485
QY 502 ACAAATCCACCCACAGATCTAGACTCAGGTCACTCTCTGAAGCCCTCGTTTCAGGGTCGATA 561
DB 486 ACAAATCCACCCACAGATCTAGACTCAGGGCACTCTCTGAAGCCCTCGTTTCAGGGTCGATA 545
QY 562 TCCCAACATCAAGGCGGTGAGGATCTTAAAGACCACTCAACAGAAAGTAAAAAAATGGA 621
DB 546 TCCCAACATCAAGGCGGTGAGGATCTTAAAGACCACTCAACAGAAAGTAAAAAAATGGA 605
QY 622 AATTTGCTAGAGAAATCCAGGCATGAAGTGAAGAAATCAGAAATCAGTGAAGAAACACAGA 681
DB 606 AATTTGCTAGAGAAATCCAGGCATGAAGTGAAGAAATCAGAGATCAGTGAAGAAACACAGA 665
QY 682 TGGTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAAGATGATTTGA 741
DB 666 TGGTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAACAGGCTTAAAGATGATTTGA 725
QY 742 GAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAAGTGAAGTGAAG 801
DB 726 GAAAGGTGAACCAACTCAAACTAAGATTCTCCGGGCCCAAGCCGAAAGTGAAGTGAAG 785
QY 802 GAAGATCTCTGAAACACAGCTATTCTTAGATGACCTGGAAATAGGCCAGGTCAGTTGTC 861
DB 786 GAAGATCTCTGAAACACAGCTATTCTTAGATGACCTGGAAATAGGCCAGGTCAGTTGTC 845
QY 862 ATCTTCTACATTGACTCGGAGAAATAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 921
DB 846 ATCTTCTACATTGACTCGGAGAAATAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 905
QY 922 CTCAGAAACCTCTATAAAGGATCGAATGCGCAAGTACCAGGAGCTGTGTCACAAACAAAG 981
DB 906 CTCAGAAACCTCTATAAAGGATCGAATGCGCAAGTACCAGGAGCTGTGTCACAAACAAAG 965
QY 982 CAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGGCGGAAATCAAAATTCATAA 1041
DB 966 CAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGTGGCGGAAATCAAAATTCATAA 1025
QY 1042 AATGAGCAAAAGGAGATGTGCCCCCAGGTCCTGAGGTCTGATCACCCTTCAGGAAGG 1101
DB 1026 AATGAGCAAAAGGAGATGTGCCCCCAGGTCCTGAGGTCTGATCACCCTTCAGGAAGG 1085
QY 1102 GAAAAGATTTCTGCAAAATGAGATAGCTCGGAGTCCGCTTCCACCCCTGCGGAAGATGA 1161
DB 1086 GAAAAGATTTCTGCAAAATGAGATAGCTCGGAGTCCGCTTCCACCCCTGCGGAAGATGA 1145
QY 1162 CTCCTGACTCCCAGGTTAAGAGTGAAGTTCACAGGCTGTCCATCCCAAGCCACTAAG 1221
DB 1146 CTCCTGACTCCCAGGTTAAGAGTGAAGTTCACAGGCTGTCCATCCCAAGCCACTAAG 1205

QY 1222 TCCAGATTCCAGAGCCTCCAGTCTTTCTGAAAGTCTCTCCCTCCCAAGCAATGAAGAAGTT 1281
DB 1206 TCCAGATTCCAGAGCCTCCAGTCTTTCTGAAAGTCTCTCCCTCCCAAGCAATGAAGAAGTT 1265
QY 1282 TCAGGCACCTGCAAGAGACCTGCGTGGAAATGTCAGAGACAGTCTTATCCAAATGGAGCG 1341
DB 1266 TCAGGCACCTGCAAGAGACCTGCGTGGAAATGTCAGAGACAGTCTTATCCAAATGGAGCG 1325
QY 1342 TCTCTTGGCCCAACAGCAGAGTGTTCACATCAGCTGCTCCGTTGCTCTTATTTGCAACAA 1401
DB 1326 TCTCTTGGCCCAACAGCAGAGTGTTCACATCAGCTGCTCCGTTGCTCTTATTTGCAACAA 1385
QY 1402 CAAACTCAGTCTAGGACATATGATCTTTACATGGAAGATCTATTGTAAGCCTCACTT 1461
DB 1386 CAAACTCAGTCTAGGACATATGATCTTTACATGGAAGATCTATTGTAAGCCTCACTT 1445
QY 1462 CAATCAACTCTTTAAATCTTAAGGCAACTATGATGAAGGCTTTGGGCACAGACCAACAA 1521
DB 1446 CAATCAACTCTTTAAATCTTAAGGCAACTATGATGAAGGCTTTGGGCACAGACCAACAA 1505
QY 1522 GGATCTATGGGCAAGCAAAATGAAACGAAGAGATTTTGGAGAGACCCAGCCAGTTGC 1581
DB 1506 GGATCTATGGGCAAGCAAAATGAAACGAAGAGATTTTGGAGAGACCCAGCCAGTTGA 1565
QY 1582 AAATGCAAGGAGAGACCCCTCACAGCCAGGGGTAGAAGA-TGCCCTTATTGCTAAGGTGG 1640
DB 1566 AAATCAAGGAGAGACCCCTCAAAAGCCAGGGGGTGAAGA-TGCCCTTATTGCTAAGGTGG 1625
QY 1641 GTGTCTCTGGTGAAGATATGGAAGCCAAAGCCCTCTCTCAGCAGAGAGAGAGACAAGC 1700
DB 1626 TTGTCTCTGGTGAAGATATGGAAGCCAAAGCCCTCTCTCAGCAGAGAGAGAGACAAGC 1685
QY 1701 CAGCTGAAACCAAGAGCTGAGG 1723
DB 1686 CAGCTGGAACCCGAACTTAGGG 1708

RESULT 14
AAH18125
ID AAH18125 standard; cDNA; 2207 BP.
XX
AC AAH18125;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:17991.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0249036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -
XX Claim 8; SEQ ID 17991; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesising polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
represent oligonucleotides, all of which are used in the exemplification
of the present invention.
XX
SQ Sequence 2207 BP; 692 A; 438 C; 512 G; 565 T; 0 other;

Query Match 42.3%; Score 1568.4; DB 22; Length 2207;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1569; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1001 AATGAGCTGAAAGCCAGTGGTGGCCAAATCAAAATTCATAAAATGGACAAAGAGAAAT 1060
DB 638 AATGAGCTGAAAGCCAGTGGTGGCCAAATCAAAATTCATAAAATGGACAAAGAGAAAT 697
QY 1061 GTGCCCCCAGGTCTCTGAGTCTGCATCACCATCAGGAAGGGGAAAGATTTCTGCAAAAT 1120
DI 698 GTGCCCCCAGGTCTCTGAGTCTGCATCACCATCAGGAAGGGGAAAGATTTCTGCAAAAT 757
QY 1121 GAGATAGCCTGGCAGTCCGTTCCACCCCTCCGGAAGATGACTCCCGTACTCCAGGTT 1180
DB 758 GAGATAGCCTGGCAGTCCGTTCCACCCCTCCGGAAGATGACTCCCGTACTCCAGGTT 817
QY 1181 AAGAGTGAGGTTCAACAGCCTGTCTATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 1240
DB 818 AAGAGTGAGGTTCAACAGCCTGTCTATCCCAAGCCACTAAGTCCAGATTCAGAGCCTCC 877
QY 1241 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 1300
DB 878 AGTCTTTCTGAAAGTTCTCTCCCAAGCAATGAAGAAGTTTCAGGCACCTGCAAGAGAG 937
QY 1301 ACCTGCGTGGAAATGTCAGAGACAGTCTTATCCAATGGAGCGTCTTTGGCCCAACCCAGCAG 1360
DB 938 ACCTGCGTGGAAATGTCAGAGACAGTCTTATCCAATGGAGCGTCTTTGGCCCAACCCAGCAG 997
QY 1361 GTGTTTTCACATCAGCTGCTTCCGTTGCTCTTATTTGCAACCAAACTCAGTCTAGGAACA 1420
DB 998 GTGTTTTCACATCAGCTGCTTCCGTTGCTCTTATTTGCAACCAAACTCAGTCTAGGAACA 1057
QY 1421 TATGATCTTTACATGGAAGATCTTATTTGAAGCCTCACTTCAATCAACTTTTAAATCT 1480
DB 1058 TATGATCTTTACATGGAAGATCTTATTTGAAGCCTCACTTCAATCAACTTTTAAATCT 1117
QY 1481 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACACACAGAGGATCTTATGGCAGGAAA 1540
DB 1118 AAGGGCAACTATGATGAAGGCTTTGGGCACAGACACACAGAGGATCTTATGGCAGGAAA 1177
QY 1541 AATGAAGCAAGAGATTTTGGAGAGACCCAGCCTTGCAAATGCAAGGAGACCCCT 1600
DB 1178 AATGAAGCAAGAGATTTTGGAGAGACCCAGCCTTGCAAATGCAAGGAGACCCCT 1237

Qy 1361 GTGTTTCACATCAGCTGCTCCGTTGCTCCTATTGCAACAACAACTCAGTCTAGGAACA 1420
Db GTGTTTCACATCAGCTGCTCCGTTGCTCCTATTGCAACAACAACTCAGTCTAGGAACA 546
Qy 1421 TATGCATCTTTACATGGAAGAATCTATTGTGAAGCCTCAGTTCGAATCAACTCTTTAAATCT 1480
Db TATGCATCTTTACATGGAAGAATCTATTGTGAAGCCTCAGTTCGAATCAACTCTTTAAATCT 606
Qy 1481 AAGGCAACTATGATGAAGCCTTTGGGCACAGACCACCAAGGATCTATGGGCAAGCAAA 1540
Db AAGGCAACTATGATGAAGCCTTTGGGCACAGACCACCAAGGATCTATGGGCAAGCAAA 666
Qy 1541 AATGAAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTGCAAAATGCAAGGGAGACCCCT 1600
Db AATGAAAACGAAGAGATTTTGGAGAGACCAGCCAGCTTGCAAAATGCAAGGGAGACCCCT 726
Qy 1601 CACAGCCCAAGGGTAGAAGATGCCCTATTGCTAAGGTGGGTGTCTCTGGCTGCAAGTATG 1660
Db CACAGCCCAAGGGTAGAAGATGCCCTATTGCTAAGGTGGGTGTCTCTGGCTGCAAGTATG 786
Qy 1661 GAAGCCAAAGGCTCTCTCAGCAGGAGAAGAGACAAGCCAGCTGAAACCAAGAAGCTG 1720
Db GAAGCCAAAGGCTCTCTCAGCAGGAGAAGAGACAAGCCAGCTGAAACCAAGAAGCTG 846
Qy 1721 AGGATCCCTGGCCACCCCTGAACTTGAAGTTTCGGAAGTGCCTTGGAGGAAGGG 1780
Db AGGATCCCTGGCCACCCCTGAACTTGAAGTTTCGGAAGTGCCTTGGAGGAAGGG 906
Qy 1781 ATCAAAAATGTCAAAGCCCAATGGCCCTCTGAAAGACGAAATCAGCAAGCCCGAAGTTCT 1840
Db ATCAAAAATGTCAAAGCCCAATGGCCCTCTGAAAGACGAAATCAGCAAGCCCGAAGTTCT 966
Qy 1841 GAGGATGTCGATCTAGATCTGAAGAAGCTAAGACGATCTTCTTCACTGAAGGAAGAAGC 1900
Db GAGGATGTCGATCTAGATCTGAAGAAGCTAAGACGATCTTCTTCACTGAAGGAAGAAGC 1026
Qy 1901 CGCCCATTCAGTGTAGCAGCTTCATTTCAAGCACCTCTGTCAAGAGCCCAAAAACCTGTG 1960
Db CGCCCATTCAGTGTAGCAGCTTCATTTCAAGCACCTCTGTCAAGAGCCCAAAAACCTGTG 1086
Qy 1961 TCCCCACCTATCAGGAAGGCTGGAGCATGTCAAGCAGAGTGAAGAGTCTGTGGGTGGA 2020
Db TCCCCACCTATCAGGAAGGCTGGAGCATGTCAAGCAGAGTGAAGAGTCTGTGGGTGGA 1146
Qy 2021 AGAGTTGCAAGAAAGGAACAAAGTGGAAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTG 2080
Db AGAGTTGCAAGAAAGGAACAAAGTGGAAAATGCCAAGGCTTCTAAGAAAGATGGGAATGTG 1206
Qy 2081 GGAAGAACACCTGGGCAACAAAGAAATCTAAGAGGAGACAGGGGAAGAGAGTAAGGAA 2140
Db GGAAGAACACCTGGGCAACAAAGAAATCTAAGAGGAGACAGGGGAAGAGAGTAAGGAA 1266
Qy 2141 GGTCTAGTTTGGAGATGGAGATGAGAAATCTTTGTAGAAAATGGTGCAGACTCCGATGAA 2200
Db GGTCTAGTTTGGAGATGGAGATGAGAAATCTTTGTAGAAAATGGTGCAGACTCCGATGAA 1326
Qy 2201 GATGATAACAGCTTCTCAAAACAAATCTCCACAAGAACCCCAAGTCTCTGAATGGTGG 2260
Db GATGATAACAGCTTCTCAAAACAAATCTCCACAAGAACCCCAAGTCTCTGAATGGTGG 1386
Qy 2261 AGTTTGTAGACAACACCTTTGCTGAAGAAATTCAGTCTCAGAAATCAGAAATCCAGGAT 2320
Db AGTTTGTAGACAACACCTTTGCTGAAGAAATTCAGTCTCAGAAATCAGAAATCCAGGAT 1446
Qy 2321 GTGGAACTCTGGGAGGAGAGTGGTCAAGAGCTCTCTGTGGAGAAACAGATAAGAGA 2380
Db GTGGAACTCTGGGAGGAGAGTGGTCAAGAGCTCTCTGTGGAGAAACAGATAAGAGA 1506
Qy 2381 AATCGGTATTATGATGAGGATGAGGATGAAGAGTGAAGAAATTCGAATGATGCTGGCCCT 2440
Db AATCGGTATTATGATGAGGATGAGGATGAAGAGTGAAGAAATTCGAATGATGCTGGCCCT 1566
Qy 2441 A 2441

Db 1567 A 1567

Search completed: January 6, 2004, 10:12:20
Job time : 910 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2004, 09:52:31 ; Search time 7186 Seconds
(without alignments)
12531.044 Million cell updates/sec

Title: US-09-890-549-16

Perfect score: 3705

Sequence: 1 ggcgcaggagcagtaggtg.....atttgaaaaaaaaaaaaa 3705

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em_estba:*

2: em_estim:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1557.4	42.0	3242	11 AK049350	AK049350 Mus muscu
2	1403.2	37.9	3108	11 AK031698	AK031698 Mus muscu
3	1401.6	37.8	3100	11 AK028186	AK028186 Mus muscu
4	1400	37.8	3160	11 AK085065	AK085065 Mus muscu

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

RESULT 1

AK049350

LOCUS

DEFINITION

AK049350

AK049350

AK049350.1 GI:26340071

HTC; CAP trapper.

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

PUBMED

10349636

REFERENCE

2

REFERENCE

1

REFERENCE

1

REFERENCE

1

REFERENCE

1

AK049350 3242 bp mRNA linear HTC 05-DEC-2002
Mus musculus ES cells cDNA, RIKEN full-length enriched library, full
clone: C30026L07 product: epithelial protein lost in neoplasm, full
insert sequence.

AK049350

AK049350

AK049350.1 GI:26340071

HTC; CAP trapper.

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

PUBMED

10349636

REFERENCE

2

REFERENCE

1

REFERENCE

1

REFERENCE

1

REFERENCE

ALIGNMENTS

949.6	25.6	1014	9	AL578990	AL578990
927.6	25.0	1045	9	AL556858	AL556858
898.2	24.2	1075	12	BM909158	BM909158
806.8	21.8	880	12	BI871114	BI871114
802.4	21.7	884	14	CD171644	AGENCOURT
787.8	21.3	1142	12	BM914155	AGENCOURT
775.4	20.9	810	12	BM721370	UI-E-E01-
774	20.9	785	10	BG740815	602633701
766.4	20.7	777	28	AQ314676	603111-10
766	20.7	828	12	BI457843	603198215
764.2	20.6	781	10	BG676749	603233-EN1
744	20.1	759	12	BM977831	602622311
732.4	19.8	759	14	CA447584	UI-H-E10-
731.8	19.8	767	12	BM981550	UI-CF-EN1
731.4	19.7	745	14	CB055210	NISC-9m08
728	19.6	768	10	BG476496	602522032
726.8	19.6	989	12	BM552304	AGENCOURT
724.8	19.6	728	13	BG631767	UI-H-F10-
722.8	19.5	839	10	BG697786	602661275
721	19.5	971	10	BE410108	601302288
719.4	19.4	725	13	BU740444	UI-E-E01-
717.8	19.4	724	12	BM980979	UI-CF-EN1
713.2	19.2	742	12	BI766182	601052810
711.2	19.2	855	12	BG777288	602664524
707.8	19.1	787	13	BU861039	AGENCOURT
707	19.1	711	14	CA447848	UI-H-E10-
706	19.1	931	12	BM449793	AGENCOURT
703.2	19.0	918	13	BU850203	AGENCOURT
698.4	18.9	719	14	CA412726	UI-H-E20-
697.8	18.8	745	10	BG698125	602660017
697.2	18.8	724	14	CB055211	NISC-9m08
696	18.8	847	13	BU603212	AGENCOURT
694	18.7	843	9	AA909018	601105-s
692.6	18.7	951	10	BF528886	602043349
692	18.7	715	9	AL048161	60285056K
687.6	18.6	770	12	BI089829	60285056K
685.4	18.5	695	14	CA420719	UI-H-FG0-
685.2	18.5	701	14	CD365699	UI-H-ET2-
683.4	18.4	752	10	BE897604	601432448
683	18.4	767	14	CB988778	AGENCOURT
682.8	18.4	1120	10	BG120125	602353491

REFERENCE
AUTHORS

3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076961
4

TITLE

JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS

Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamataka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
Quackenbush,J., Schrim,L.M., Staubli,F., Suzuki,R., Tonita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Machionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
Wyshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.,
and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

TITLE

JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3242)

TITLE

JOURNAL
REFERENCE
AUTHORS

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hangaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihata,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saich,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222,
Fax:81-45-503-9216)

TITLE

JOURNAL

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues
Please visit our web site for further details.
URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>
Location/Qualifiers

COMMENT

FEATURES

source	1. .3242	961 a	797 c	863 g	621 t
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/strain="C57BL/6J"				
	/db_xref="FANTOM DB:C330026L07"				
	/db_xref="taxon:10090"				
	/clone="C330026L07"				
	/cell_type="ES cells"				
	/clone_lib="RIKEN full-length enriched mouse cDNA library"				
	105..2386				
CDS	/notes="unnamed protein product; epithelial protein lost in neoplasm (MGD MG1:1920992, GB NM_023063, evidence: BLASTN, 98%, match=3086) putative"				
	/codon_start=1				
	/protein_id="BAC13699.1"				
	/db_xref="GI:26340072"				
	/translation="MESTFENRRQMTLSLRVTAKELSLVKNKNSAIVEIFSXYOKA AEEANERKNNPESLPOHFRGLTSLVKKMVENPVAGAEFHTDSI.PNSSSGGHTAD YPAEYTDPAQGVADRREHTQPRFGRSEAVIQSRYPRSEHNDPRAQTESQK MENCLGDSRHEAEKPDGTETSGKIEKYNVPLNRLKMKFKEGHNQTKSLUMTSRN AGRRUSENNCKLDDWEICAGHLSSAFNKSNEKRNLEPLRLSETSKIDRMAYQA AVSKQSPASYTNELTSKTHKEQKNVPPPEAYSVHOEGSKVSTENSLVALS VPAEDDTCSQVKSBAQPMHPKPLSPDARTSSLPESSPSKTKYFOAKESQVCO KTVYPMERLLANQQPHISCFSCYCNKLSLGTYSLSHGRYICKYFHNQLFKSGNY DSGFGRKHQKLDKPSHNEETLGRPAQPPNAGESPHSPGVEDAPIAKYGVLAASMEA KASSQREDKPAETKRLRIAWPPPAELGGSGSALLEGIKVSKPKWPPDDVCKTEAP EVDVLDKLRSSSLKRSRFTVAASRTSSISKPKASSPSLRKWSSESQSEEFQ GGIATMERKOTENARPSGKENVKSRWQGVPRSKDRSSFELESENEMENGANTAE DNHVAHQOSPLEPEAPGSGFVDITAAKEFTIQKSDQVGFWEGEVVRVLSVERQI KNRYVDEDEDEE"				
BASE COUNT	961 a	797 c	863 g	621 t	
ORIGIN					
	Query Match	42.0%;	Score 1557.4;	DB 11;	Length 3242;
	Best Local Similarity	78.3%;	Pred. No. 9.4e-191;		
	Matches 1956;	Conservative	0;	Mismatches 506;	Indels 35; Gaps 6;
QY	111	CAGGTATTTTCAGTGTCTGTAGACAAGATGGAATCATCTCCATTTAATAGACGGCAATGGA	170		
Db	79	CAGGTGTCTCAGAGTCTGTAGACAAGATGGAATCGACTCCATTTAATAGACGGCAATGGA	138		
QY	171	CCTCATTATTCAGTGTGTAACAGCCAAAGAACTTTCTTCTTCTCAACAGAAACAAGTGTAT	230		
Db	139	CTTCCCTGTCTATGAGAGTAACAGCCAAAGAGCTTTCTCTTGTCAACAGAAACAAGTGTAT	198		
QY	231	CGGCTATTCTGGAAATATTCTCCAGTACCAGAAAGCAGCTCAAGAAACAACAATGAGAGA	290		
Db	199	CGCAATTTGTGAAATATTCTCAGAGTACCAGAAAGGCTGCTGAAGAAGCCAAACATGGAAA	258		
QY	291	AGAAGAGAAAGTAAACACCGGAAATCTCTCCAGCAGCTTTAGAAAGGGGACCCCTGACTGTGT	350		
Db	259	GGAGAGAAATTAACCTGAAAGTCTGCCCCAGCAGCTTTAGAGAGAGGACCCCTGTCTGTGT	318		
QY	351	TAAAGAAAGTGGAGAAACCCAGGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGA	410		
Db	319	TAAAGAAAGTGGAGAAACCCGGTGGTGGGGCAGAAATTCACACAGACTCCTGCGCA	378		
QY	411	ACAGCAGACTCAGATTAGGCAACAGAGACCATCTCTCTCTCTGCTGAAGTGAACAACCCAG	470		
Db	379	ACAGCAGCAGTGGAGGTGGGCAACAGCGGAGTACCCCTCTGCTGAGGAGCAACCCAGCAAGC	438		
QY	471	CTGCTTCTGGAGCCAAAGCTGACCAAGAGAAACAATCCACCCAGACTAGACTCAGGT	530		
Db	439	CTGCTCTCTGGAGTCAAGCTGACCGGGAAGAGCACCACCGCCCAAGCTAGATTGGAT	498		
QY	531	CACCTCTCTGAAGCCCTCTCTCAGGCTCGATATCTCCCAATCAAGGACCGGTGAGGATCTTA	590		
Db	499	CTCGTCCGGAAGCAGCTTATCCAGACCCGGTATCTCTCGCTCAGAGAAACACCCAGATTTA	558		
QY	591	AGACCACTCAACAGAAAGTAAAAAATTTGTTAGGAGAAATTCAGGAGATTCACCGCATGAG	650		
Db	559	AAGCCCGGCGCCAGGAGAGCCAAAAAATGGAATACTGTCTGGGAGATTCACCGCATGAG	618		

QY 651 TAGAAAAATCAGATAAGTAAACACAGATGCTTCGGGCAAAATAGAGAAATATATG 710
Db 619 CAGAGAGCCAGAGACGGGCGAAACACAGAAATTCAGGCAAAATAGAGAAATACACG 678
QY 711 TTCGCTGAACAGAGCTTAAGATGATGTTTGAGAAAGGTGAACCAACTCAAACTAAGATT 770
Db 679 TTCACCTGAATAGACTGAAGATGATGTTTGAGAAAGGTGAACCAACCAACCAAGAGTC 738
QY 771 TCCGGGCCCAAGCCGGAAGTGAAGTGAAGGAAGATCTCTGAAACAGCTATTCTCTAG 830
Db 739 TCTGGACCCCAAGCCGGAATGCGGGTGAAGAGGCTCTCTGAAACCAACTGTTCCCTGG 798
QY 831 ATGACCTTGGAAATAGGCCAGGTGAGTTGTCATCTTCTACATTTGACTCGAGAAATG 890
Db 799 ATGACTGGGAAATAGGTGCGGTATTGTCTTCTCTGCAATCAACTCGGAGAAACG 858
QY 891 AGAGTAGACGAAATCTTGGAACTTCCAGCCCTCTCAGAAACCTCTATAAGGATCGAATGG 950
Db 859 AGAGTAAGAGAAATCTGAGCTGCCAGCCTCTCAGAAACCTCCATAAAGGACCGCATGG 918
QY 951 CCAAGTACCAAGGAGCTGTGTCCAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGA 1010
Db 919 CCAAGTACCAAGGCTGCAAGTGTCCAAAGCAGAGCAGCCAGCCAGCTACACAAATGAGCTGA 978
QY 1011 AAGCCAGTGTGGCGAAATCAAAATTCATAAATGGAGCAAAAGGAGAAATGTGCCCCAG 1070
Db 979 AAACCACT-----GAAAGCAAACTCATAAATGGGAAACAGAAAGGAGAAATGTGCCCCAG 1032
QY 1071 GTCTGAGGCTGTCATCACCCATCAGAAAGGGGAAAGATTTCTGCAAAATGAGAAATAGC 1130
Db 1033 GTCCGAGGCTACAGGCTCCATCAGGAAGGAGCAAGGTTTCTACAACTGAGAAATAGTC 1092
QY 1131 TGGAGTCCGTTTCACCCCTGCCGAGATGACTCCCGTGACTCCAGGTTAAGAGTGAGG 1190
Db 1093 TAGTGGCCCTCTGTCTCCCTGCTGAAGATGACACCTGTAACTCCAGGTGAAGAGCGAGG 1152
QY 1191 TTCAAGAGCTGTCCATCCCAAGCCACTAAGTCCAGATTCAGAGGCTCCAGCTTTCTG 1250
Db 1153 CCCAGAGGCTATGCACTTAAGCCGTGAGTCTGACGCCAGGACCTCCAGTCTCCCG 1212
QY 1251 AAAGTTCTCTCCCAAGCAATCAAGAGTTTCAGGCACTGCAAGAGAGACCTCGGTGG 1310
Db 1213 AAAGTTCTCTCTCCAAACAGCGAAAGTTTCAGGCGCGGCAAGGAGAGCTGCGTGG 1272
QY 1311 AATGTCAGAAAGAGCTATATCCAAATGGAGGCTCTTGGCCAAACAGCAGGTGTTTCACA 1370
Db 1273 AGTGCAGAGAGCGGTACCCCATGGAACGGCTCTGGCCAAACAGCAGGTGTTTCACA 1332
QY 1371 TCAGCTGTTCCGTTGCTCTATTGCAACAACTCAGTCTAGGAACATATGCAATCT 1430
Db 1333 TCAGCTGTTCCGATGCTCTACTGCAACAACTCAGTCTAGGAACATATGCAATCT 1392
QY 1431 TACATGGAAGATCTATTGTAAGCCTCACTTCAATCAACTCTTTAAATCTAAGGCAACT 1490
Db 1393 TACATGGAAGATCTACTGTAAGCCTCACTTCAATCAACTCTTTCAAGTCCAAAGGCAACT 1452
QY 1491 ATGATGAAGCTTTGGGCAGACACCAAGATCTATGGCCAGCAAAATGAAAGC 1550
Db 1453 ATGACAGAGGCTTTGGGCACAGCACAAGATCTGTGGCAAGCAAGAGCGCAATG 1512
QY 1551 AAGAGATTTTGGAGACACCGCCAGCTTGCAAATGCAAGGAGAGACCCCTCACAGCCAG 1610
Db 1513 AGGAGACTTTGGGAGACACAGCCAGCCGCTTAATGACGGGAGAGCCCTCATAGCCCG 1572
QY 1611 GGTAGAAAGATGCCCTATTGCTAAGTGGGTGCTCTGCTGCTCAAGTATGGAAGCAAGG 1670
Db 1573 GTGTAGAAGATGCCCCATCGCAAGGTGCGGTGCTGCGCGCAAGTATGGAAGCAAGG 1632
QY 1671 CTTCTCTCAGCAGGAGAGGAGCAAGCCAGCTGAACCAAGAGAGCTGAGGATCGGCT 1730
Db 1633 CTTCTCTCAGAGGAGAGGAGGATGAGCCTGCGAGACCAAGAGCTGAGGATCGGCT 1692

QY 1731 GGCCACCCCCACTGAACTTGGAAATTCAGGAAGTGCCTTGGAGGAAGGGATCAAAATGT 1790
Db 1693 GGCTTCGGCCAGCCGAGCTGGCGGTTCGGGAAGTGCCTTGGAGGAAGGGATCAAAATGT 1752
QY 1791 CAAAGCCCAATGGCTCTCTGAAGACGAATCAGCAAGCCCAAGTTCCTCTGAGGATGTCTG 1850
Db 1753 CGAAGCCCAATGGCTCCGAGGATGACGCTCTGCAAGCAGGAGGCCCCCGAGGATGTAG 1812
QY 1851 ATCTAGATCTGAAGAGCTTAAGACGATCTTCTCACTGAAGGAAAGAACCCGCCCTTCA 1910
Db 1813 ATCTGACCTGAAGAAGCTACGCGGTCTTCGCTCTGAAGGAGGAGGCCCCCGCTTTA 1872
QY 1911 CTGTAGCAGCTTCAITTTAAAGCACCTCTGTCAAGAGGCCCAAAAACCTGTGTCCACCTA 1970
Db 1873 CGGTAGCAGCTTCTTTTCAACCTCTTCCATCAAGAGCCCAAGGCCCTCTTCGCATCTC 1932
QY 1971 TCAGAAAGGCTTGGAGCATGTCTAGACGAGTGAAGAGTCTGTGGGTGGAAGATTGCA - 2029
Db 1933 TCAGAAAGGCTTGGAGCAGTCCGAGCAGTGAAGAGTTTGGAGGAGGGATAGCGACA 1992
QY 2030 --GAAAGGAAAACAAGTGGAAAATGCCAAGCTTCTAAGAAGAAATGGGAATGTGGGAAAA 2087
Db 1993 TGAAGGAAAACAAGGAAAATGCCAGACCTCTTGGGAGAAAGAGAAATGTGGGAAAA 2052
QY 2088 CAACCTGSCAAAACAAGAAATCTAAAGGAGAGACAGGGAAGAGAAATGAAGAAAGTCTATA 2147
Db 2053 CACGCTGGC-----AGGCGAGGAAGTCCCGAGGAGCAAGGACCTAGTT 2097
QY 2148 GTTTCGAGATGAGAAATCAGAAATCTTTAGAAAATGGTGCAGACTCCGATGAAGATGATA 2207
Db 2098 CTTTTCGAGCTGGAGAGTGAAGATTTTATGAAAATGGAGCAACATAGCTGAAGATGACA 2157
QY 2208 ACAGCTTCTCAACAAACAATCTCCACAGAACCCCAAGTCTCTGAATTTGGTCGAGTTTG 2267
Db 2158 ACCATGTCACGCAAGCAGTCCCGTTAGAACCCGAGGCTCCGGGTTGGTCTGGCTTTG 2217
QY 2268 TAGACAACACCTTTCTCTGAAGAAATTCACCTACTCAGAAATCCAGAAATCCCGAGGATGGAAC 2327
Db 2218 TAGACAACACCGCTGCTTAAGAAATTCATACCCAGAAATCAAAAATCCCAAGATGCGGT 2277
QY 2328 TCTGGAGGGAGAGTGTGTCAAGAGCTCTCTGTGGAAGAACAGATAAAGAGAAATCCGT 2387
Db 2278 TCTGGAGGGAGAGTGTGTCCGAGAGCTGTCTGTGAGGAACAGATTAAGAGAAACCGT 2337
QY 2388 ATTATGATGAGATGAGATGAAGAGTGAAGAAATTTGCAATGATCTGGGCTTTAAATTC 2447
Db 2338 ACTATGATGAGAGCAGGATGAAGAAATGACCGGCCCCCATGGTTCTGCCC-----TTA 2390
QY 2448 GTTATGTTAGCGAGCCACTGCCCTTTGTCAAAATGTGATGCCACATAGCAGGATCCCC 2507
Db 2391 GATTTTGGTTAGCAACTCACTGCCCCCTTTATCAA---GTGGTGGCAGAAACAGCAGCTT 2447
QY 2508 AGCATGAAATGTAATTTACTTTGGAAGTAACTTTGGAAAAGAAATTCCTTCT-TAAATCAA 2566
Db 2448 GGCCCAAGGTCATTGATGTGGAGCACTTTGGAAATGAAATTTCCCGCTAGCAAAACAG 2507
QY 2567 AAACAAAACAAAACCAACAAAACACATTTCTAAAT 2603
Db 2508 GGCAAAAACAAAGCAAAAACAAAACAGACAAACAACT 2544

RESULT 2

AK031698

LOCUS

DEFINITION

AK031698 3108 bp mRNA linear HTC 05-DEC-2002
Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
enriched library, clone:6030492D07 product:epithelial protein lost
in neoplasm, full insert sequence.

ACCESSION

AK031698.1 GI:26327552

VERSION

HTC; CAP trapper.

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

ORGANISM	Mus musculus	TITLE	Direct Submission
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp). URL: http://genome-gsc.riken.go.jp/. Tel: 81-45-503-9222, Fax: 81-45-503-9216)
AUTHORS	Carninci, P. and Hayashizaki, Y.	COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
TITLE	High-efficiency full-length cDNA cloning		Please visit our web site for further details.
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)	URL: http://genome-gsc.riken.go.jp/	URL: http://fantom-gsc.riken.go.jp/
MEDLINE	9279253	Location/Qualifiers	
PUBMED	10349636	1. 3100	
REFERENCE		/organism="Mus musculus"	
AUTHORS		/mol_type="mRNA"	
		/strain="C57BL/6J"	
		/db_xref="FANTOM_DB:2610034J22"	
		/db_xref="taxon:10090"	
		/clone="3610034J22"	
		/tissue type="whole body"	
		/clone_lib="RIKEN full-length enriched mouse cDNA library"	
		/dev_stage="10 days embryo"	
		448..2229	
		/notes="unnamed protein product; epithelial protein lost in neoplasm (MGD MG1:1920992, GB NM_023063, evidence: BLASTN, 99%, match=3086)	
		putative"	
		/codon_start=1	
		/protein_id="BAC25798.1"	
		/db_xref="GI:26389834"	
		/translation="MENCIGDSRHAERKPESTENTSGKIEKYNVPLNPLKMPFEK	
		SEINQKSLWTOGRNAGGRRLSENCSLDDEICAGHLSASRNSEKSKRLELRL	
		SETSIKDRMAKYCAAVSKSSPASVTNELKTESKTHWEKENVPPGPEACSVHOEG	
		SKVSTIENSLVSPAEEDTNSQVSEAOQPMHPKLPUSDPARTSSLPESPSTAK	
		KFOAPAKESCVEQKTVYPMERLLANQOVFIHSCFSCYCNKLSLGTVASLHGRTYC	
		KPHNFQFKSGNYDEFGHKHKLWASKSDNEETLGRPAOPPNAGSPHSPGVEDA	
		PIAKGVLAASMEAKASSOREDKPAETKRLRIAWPPPAEIGGSGSALKEEIGKVKP	
		KWPPDDVCKTEAPEDVDLKLKRSLSLKERSPFVAASFRTSSIKSPKASPSL	
		RKGNSESESEFPGGIATMERKOTENARPSEKENVKSRWQGEVPRSKDRSPFL	
		ESENFMENAGIAEDDNDVHVAQSPLEAPGWSGFVDTAAKEFTTHNOKSQDVGFV	
		EGEVVRELSVEEQIKENRYDEDEEE"	
		BASE COUNT 908 a 759 c 824 g 609 t	
		ORIGIN	
		Query Match 37.8%; Score 1401.6; DB 11; Length 3100;	
		Best Local Similarity 77.5%; Pred. No. 9.3e-171;	
		Matches 1787; Conservative 0; Mismatches 484; Indels 35; Gaps 6;	
		QY 302 AACACCGAAATCTCTCCAGCAGCTTTAGAAAGGGACCTGACTGTGTTAAAGAAAG 361	
		Db 133 AACCTGAAAGTCTGCCAGCAGCTTTAGAGAGGACCTGTCTGTGTTAAAGAAAG 192	
		QY 362 TGGGAGAACCCAGGCTGGGAGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACT 421	
		Db 193 TGGGAGAACCCGGTGGCTGGGGCAGAAATTCACAGACTCACTCCCAACAGCAGCAGT 252	
		QY 422 GAGATTAGCAGCAGAGCAGACCATCTCTCTGCTGAAGTGACAAGCCAGCGCTGCTTCTGA 481	
		Db 253 GAGGTTGGGCACACAGCGGACTTACCCTCTGCTGAAGTGACGGACAAGCCTGCTCTTGA 312	
		QY 482 GCCAAGCTGCCAAGAGACAAATCCACCCAGATCTAGACTCAGGTCACTCTCTGTA 541	
		Db 313 GTCAGAGTGCACCGGGAAGAGACACCCAGGCCAAACCTAGATTGAGTCTCGTCCGGA 372	
		QY 542 GCCCTCGTTCAGGGTCGATATCCCCACATCAAGGACGCTGAGGATCTTAAGACACTCA 601	
		Db 373 GCAGTTATCCAGAGCCCGTATCTTCCTGCTCAGAGAACGCCACGATTTTAAAGCCCGCC 432	

QY 602 ACAGAAAGTAAATAATGGAATTTGTCTAGAGAAATCCAGGCATGAAGTAGAAAAATCA 661
Db 433 ACGGAGAGCAAAAAATGAAAAATGTCTGGAGATTCCAGGCATGAAGCAGAGAAGCCA 492
QY 662 GAAATCAGTGAATAACACACAGATGCTTCGGGCAAAATAGAGAAATATAATGTTCCGCTGAAC 721
Db 493 GAGACGAGCAAAACACAGAACTTCAGGCAAAATAGAGAAATACAAAGTTCCACTGAAT 552
QY 722 AGGCTTAAGATGATGTTTGAGAAAGGTGAACCAACTCAAACTGAAGTTCTCCGGSCCCAA 781
Db 553 AGACTGAAGATCATGTTTGAGAAAGGTGAACACACCAACCAAGAGTCTCTGAGACCCAA 612
QY 782 AGCCCAAGTGCAAGTGGAGAGAGATCTCTGAAAAACAGCTATTCTCTAGATGACCTGAA 841
Db 613 AGCCCAATGCGGGTGGAGAGGCTCTGAAAAACAACTGTTCCCTCGATGACTGGAA 672
QY 842 ATAGCCCCAGGTCAGTTGCTATCTTACATTTGACTCGGAGAAATAGAGTAGACA 901
Db 673 ATAGTGGGGTCAATTTGTTCTCTCTGATTCAACTCGAGAAAAACGAGTAAGAGG 732
QY 902 AATCTGGAATTTCCACGGCTCTCGAAAACTCTATAAAGGATCGAATGGCCCAAGTACCAG 961
Db 733 AATCTGGAGTCCACGGCTCTCAGAACTTCCATAAAGGACCGCATGGCCCAAGTACCAG 792
QY 962 GCAGCTGTGTCAAAACAAAGCAGCTCAACCAACTATACAAATGAGCTGAAAGCCAGTGGT 1021
Db 793 GCTGAGTGTCCAAAGCAGACAGCCAGCCAGCTACACAAATGAGCTGAAACCACT--- 849
QY 1022 GGCGAATCAAAATTCATAAATGGAGCAAAAGAGAGATGTGCCCCAGGTCTCTGAGGTC 1081
Db 850 ---GAAAGCAAAATCTCAATAATGGGAACAGAAAGAGAAATGTGCCCCAGGTCTCCGAGGCC 906
QY 1082 TGCATCACCCATCAGGAAGGGGAAAGATTTCTCCAAATGAGATAGACTGAGGTTCAACAGCCT 1141
Db 907 TGCAGCTCCATCAGGAAGGAGCAAGGTTTCTACAACTGAGAAATAGTCTAGTGGCCCTC 966
QY 1142 TCCACCCCTGCGGAGATGACTCCCGTACTCCAGGTTAAAGTGAAGTTCAACAGCCT 1201
Db 967 TCTGTCCCTGCTGAAGATGACACCTGTAACTCCAGGTGAAGCAGGAGCCAGCAGCCT 1026
QY 1202 GTCCATCCCAAGCCACTAAGTCCAGATCCAGAGCTCCAGTCTTTCTGAAAGTTCTCCT 1261
Db 1327 ATGACACCTTAAGCCGCTGAGTCTGAGCCAGGACCTCCAGTCTTCGGAAAGTCTCCT 1086
QY 1262 CCCAAACCAATGAGAAAGTTTCAGGCACCTGCAAGAGACCTCGGTGGAATGTCAGAAG 1321
Db 1087 TCCAAAACAGGAAGAAGTTTCAGGCGCCGCAAGAGAGAGCTGCGTGGAGTGTGAGAAG 1146
QY 1322 ACAGTCTATCCAAATGGAGGCTCTCTGGCCAAACAGCAGGTTTACATCAGCTGCTTC 1381
Db 1147 ACGGTGTACCCCATGGAAACCGCTCTGGCCAAACAGCAGGTTTACATCAGCTGTTTC 1206
QY 1382 CGTTGCTCTATTGCAACAAACAACTCAGTCTAGGAACATATGCATCTTTACATGGAAGA 1441
Db 1207 CGATGCTCTTACTGCAACAAACAGCTCAGTCTAGGAACATATGCATCTTTACATGAGCA 1266
QY 1442 ATCTATTGAAGCTCATTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGGC 1501
Db 1267 ATCTATTGAAGCTCATTCAATCAACTCTTTAAATCTTAAGGGCAACTATGATGAAGGC 1326
QY 1502 TTTGGGCACAGACACACAGGATCTATGGGCAAGCAAAATGAAACCGAAGAGATTTTG 1561
Db 1327 TTTGGGCACAGCAGGATTAAGGATCTGTGGGCAAGCAAGAGCGCAATGAGGAGACTTTG 1386
QY 1562 GAGAGACCAAGCCAGCTTCAATGCAAGGGAGACCCCTCACAGCCAGGGGTAGAGAT 1621
Db 1387 GGGAGACCAAGCCAGCCGCTTATGAGGGGAGAGCCCCCATAGCCCGGTGTAGAGAT 1446
QY 1622 GCCCTATTGCTAAGTGGGTCTCTGGCTGCAAGTGAAGCAAGCCCTCTCTCAG 1681
Db 1447 GCCCCCATCGCAAGGTGCGGTCTGTCGGGCAAGTATGGAAGCAAGGCTCTCTCAG 1506
QY 1682 CAGGAGAGGAAGACAAGCCAGCTGAACCAAGAGCTGAGGATCGCTGGCCACCCCTC 1741

Db 1507 AGGAGAGGGAGGATTAAGCCTCGGAGACCAGAAGCTGAGGATCGCCTGCCCTCCGCCA 1566
QY 1742 ACTGAACTTGGAAAGTTTCCAGGAAGTCCCTTGGAGAAAGGGATCAAAATGTCAAAGCCCAAA 1801
Db 1567 GCCAGCTGGCGGTTCCGGAAGTGCCTTGGAGAAAGGGATCAAAGTATCGAAGCCCAAG 1626
QY 1802 TGGCTTCTCTGAAGACGAAATCAGCAAGCCCGAAGTTCTCTGAGGATGTCGATCTAGATCTG 1861
Db 1627 TGGCTTCCGAGGATGACGTCTCAAGACGAGGCGCCCGGAGGATGTAGATCTGGACCTG 1686
QY 1862 AAGAAGCTAAGACGATCTTCTTCACTGAAGAAAGAGCCGCCCATCTCACTGTAGAGCT 1921
Db 1687 AAGAAGCTGCGGGGCTTCTGCTCCCTGAAGGAGAGGCGCCCGCTTTACGGTAGCAGCT 1746
QY 1922 TCATTTCAAGACACCTCTGTCAAGAGCCCAAAACCTGTGTCCCCACCTATCAGAAAGGC 1981
Db 1747 TCCTTTGAACTCTTCCATCAAGAGCCCAAGGCTCTTTCGCCATCTCTCAGAAAGGT 1806
QY 1982 TGAGCATGTTCAGAGCAGAGTGAAGAGTCTGTGGTGAAGAGTTGCA---GAAAGGAAA 2038
Db 1807 TGGAGCAGTCCGAGCAGAGTGAAGAGTTCGAGGAGGATAGCGAGGATGGAAGGAAA 1866
QY 2039 CAAGTGGAAAAATCCCAAGGCTTCTTAAGAAATGGGAATGGGAAAAACAACCTGGCAA 2098
Db 1867 CAAACCGAAAAATGCCAGACCTCTGGGAGAAAGAGAAATGTGGGAAAAATCACCTGGCAG 1926
QY 2099 AACAAAGAATCTAAAGGAGAGACAGGAGAGAGTGAAGGAAGGTCATAGTTTGGAGATG 2158
Db 1927 -----GGCAGGAAGTGCAGAGAGCAAGGACCGTAGTTCTTTTGAGCTG 1971
QY 2159 GAGAAATGAGAACTTTGTAGAAAAATGGTCAGACTCCGATGAAGATGATAACAGCTTCCTC 2218
Db 1972 GAGAGTGAAGATTTTATGAAAAATGGAGCAACATAGCTGAAGATGACAACTGTCCAC 2031
QY 2219 AAACAACAATCTCCACAGAACCCCAAGTCTCTGAATTTGGTTCGAGTTTGTAGACAACACC 2278
Db 2032 GCACAGCAGTCCCGGTAGAACCCGAGGCTCCGGTTCGGTCTGGCTTTGTACACACCAG 2091
QY 2279 TTTGCTGAGAATTCACACTACTCAGAAATCCAGGATCCGAGATGTGGAATCTCGGAGGGA 2338
Db 2092 GTGCTTAAAGAAATTCACACCCCAATCAAAATCCCAAGATGTCCGGTTCCTGGAGGGA 2151
QY 2339 GAAGTGGTCAAGAGACTCTCTGTGGAAGAACAGATTAAGAGAAATCCGTTATATGATGAG 2398
Db 2152 GAAGTGGTCCGAGAGCTGTCTGTGAGGAACAGATTAAGAGAAACCGGTACTATGATGAG 2211
QY 2399 GATGAGGATGAAGAGTGACAAATTCGAATGATGTGGCCCTTAATTCATGTTAGTGTTA 2458
Db 2212 GACGAGATGAAGATGACCGGCCCATGGTCTGGCC-----TTAGATTTTGGTTA 2264
QY 2459 GCGAGCAGTCCCTTTGTCAAAATGTGATCAGATGAAGCAGGATCCAGCATGAAATG 2518
Db 2265 GCAACTCAGTCCCTTTATCAA---GTGGTGGCAGAAACAGGAGCTTGGGCCAAGG 2321
QY 2519 TAATTTACTTGAAGTAACTTTGGAAGAAATTCCTCTCT-TAAATCAAAACAAAACAA 2577
Db 2322 TCATTCATGTGGAGCAACTTTGGAATGAATTCCTCCGTAGCAAAACAGGCAAAACAA 2381
QY 2578 AAAAAACAAAAAACACATTTCTAAAT 2603
Db 2382 GCAAAAACAAAAACAGACAAACAACT 2407

RESULT 4
AK085065
LOCUS

DEFINITION AK085065 3160 bp mRNA linear HTC 05-DEC-2002
Mus musculus 13 days embryo lung cDNA, Riken full-length enriched
library, clone:D430032103 product:epithelial protein lost in
neoplasm, full insert sequence.

ACCESSION AK085065
VERSION AK085065.1 GI:26351432
KEYWORDS HTC; CAP trapper.

Mus musculus (house mouse)	
Mus musculus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
1	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikolaic, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stauber, F., Suzuki, R., Tonita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hum, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohtsuki, S. and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,

Muramatsu, M. and Hayashizaki, Y.	
Direct Submission	
Submitted (16-APR-2002)	Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cdna library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	
Please visit our web site for further details.	
URL: http://genome-gsc.riken.go.jp/	
URL: http://fantom.gsc.riken.go.jp/	
Location/Qualifiers	
1. 3160	
/organism="Mus musculus"	
/mol_type="mRNA"	
/strain="C57BL/6J"	
/db_xref="FANTOM_DB:D430032103"	
/db_xref="taxon:10090"	
/clone="D430032103"	
/tissue type="lung"	
/clone_lib="RIKEN full-length enriched mouse cDNA library"	
/dev_stage="13 days embryo"	
501. 2282	
/note="unnamed protein product; epithelial protein lost in neoplasm (MGD MG1:1920992, GB NM_023063, evidence: BLASTN, 99%, match=3086)"	
putative"	
/codon_start=1	
/protein_id="BAC39353.1"	
/db_xref="GI:26351433"	
/translation="MENCGLGSRHEAEPKTSNTSETSGKIEKYVPLNELKWMPEKG ENOTKSLWTOSSNAGRLSENCSLDWEIGAGHLSSENFSEKNSKRNLELPLRL SETSKIDPMKYQAAYSVKQSFASVYNELKTESKTHKWEQENVPPEACSVHQEG SKVSTTENSVALSVPAEDTNSQVSEAOQHPKPLSPDARTPTSLPEASPSKTAK KFPAPKSCVECTVYEMERLANQVFIHISCFRCYCNKLSLGTVAESLHGRIYC KPFNQVLKSKYODETFGHQKDLWASNDNEETLGRPAQPPNAGSPHSQVDEDA PIAPKVLQAASMEAKASSORREDKPAETKLRJAWPPPAELGGSGSALEGIKVSKP KWPEDDVCKTEAPEDVDLKLRRSSSLKXSRPFTVAASFRTSSISPAKSPSL PKGNSSESQSEFFGGGIATMERKQENARPENVGKRWQGEVPRSKDRSSEF ESENFMENAGIAEDDHNHQAQSPLEPAEPWGSFVDTTAAKEFTTQNXSQDVGFW EGEVVRLEISVVEQIKRNYDEDEDEE"	
BASE COUNT	914 a 777 c 843 g 626 t
ORIGIN	
Query Match	37.8%; Score 1400; DB 11; Length 3160;
Best Local Similarity	77.5%; Pred. No. 1.5e-170;
Matches 1786; Conservative	0; Mismatches 485; Indels 35; Gaps 6;
QY	302 AACACCGAAATCTCTCCAGCAGCTTTTAAAGGGGACCCCTGACTGTGTTAAAGGAAG 361
Db	186 AACCTGAAAGTCTGCCCGACACTTTTAAAGAGGGACCCCTGTCTGTGTTAAAGGAAG 245
QY	362 TGGGAGAACCCAGGCGTGGGAGCAGAGTCTCACACAGACTCTCTACGGAACAGCAGCACT 421
Db	246 TGGGAGAACCCGGTGGCTGGGGCAGATTCCACACAGACTCACTGCCAAACAGCAGCAGT 305
QY	422 GAGATTAGGCACAGCAGACCATCTCTCTGCTGAAGTGACAGCCACGCTGCTTCTTGA 481
Db	306 GAGGTGGGCACACAGCGGACTACCTCTGCTGAAGTGACGCGACAGCTGCTCTTGA 365
QY	482 GCCAAAGCTGACCAAGAGACAATTCACCCAGATCTTAGACTCAGGTCACTCTCTGAA 541
Db	366 GTCAGAGCTGACCGGGAAGACACACCCAGCCCAAACTAGATTGGATCTCTGTCGGAA 425
QY	542 GCCCTCTGTCAGGTCGATATCCCCACATCAAGGACCGTGAGGATCTTAAAGACCACTCA 601
Db	426 GCAGTTATCCAGAGCCGGTATCTCTGCTCAGAACACGACCCAGCATTTTAAAGCCAGGCC 485